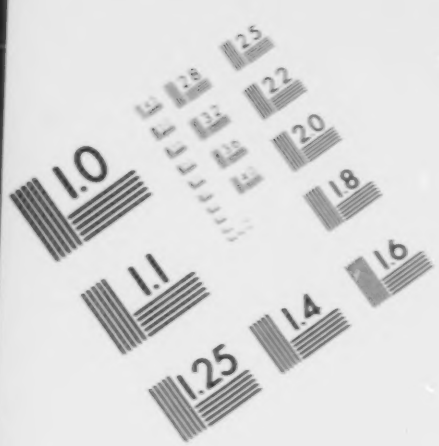
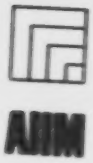


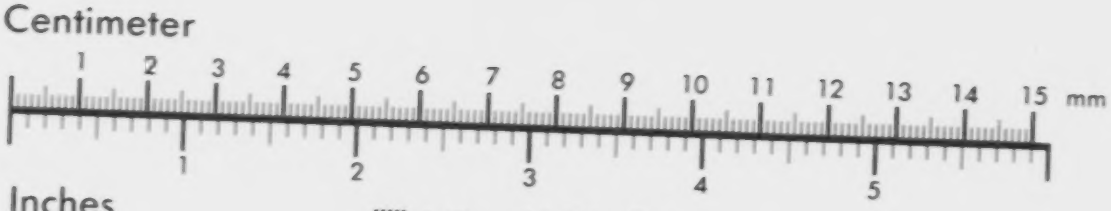
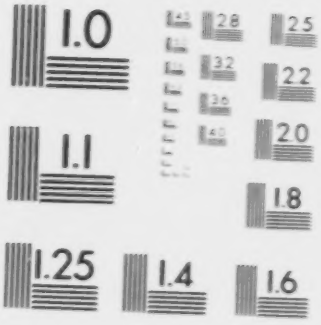


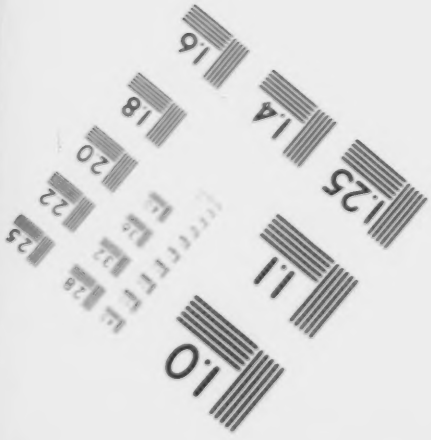
AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

301/587-8202





MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

DOCUMENT RESUME

ED 401 341 TM 026 210

AUTHOR Ballator, Nada
TITLE The NAEP Guide: A Description of the Content and Methods of the 1994 and 1996 Assessments. Revised Edition.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-97-586
PUB DATE Oct 96
NOTE 95p.; For the 1990 and 1992 Assessments, see ED 325 504.
PUB TYPE Reports - Descriptive (141) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Achievement; *Educational Assessment; Elementary School Students; Elementary Secondary Education; Geography; Grade 4; Grade 8; Grade 12; High School Students; History; Mathematics; National Surveys; Reading; *Research Methodology; Research Reports; *Sampling; Sciences; *Scoring; Surveys; Test Construction; *Test Content; Test Use
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This guide to the National Assessment of Educational Progress (NAEP) describes its purpose, contents, uses, and research methodology by presenting 21 questions about the NAEP program. The NAEP is mandated by Congress to survey the educational accomplishments of U.S. students and to monitor changes in these accomplishments. It tracks the educational achievement of 4th, 8th, and 12th grade students over time in selected content areas. Although the Commissioner of Education Statistics is responsible for NAEP operations and technical quality control, the program is governed by the National Assessment Governing Board. The assessment is implemented by the Educational Testing Service. The national main assessments in 1994 measured achievement in reading, history, and geography; in 1996, the main assessments covered mathematics and science. In both years there were smaller special assessments. Long-term assessments are also conducted to measure trends in academic achievement. In 1996, approximately 12,000 schools provided nearly 480,000 students for all components of the national and associated state samples. Samples are chosen using a multistage design involving stratification. Details about sampling and analysis are provided. Scoring and reporting are also discussed. Eight sources are listed for further reading. (Contains 6 tables and 25 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 401 341

TM

NATIONAL CENTER FOR EDUCATION STATISTICS

The NAEP Guide

A Description of the Content and Methods of the 1994 and 1996 Assessments

Revised Edition
October 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. Department of Education
Office of Educational Research and Improvement

BEST COPY AVAILABLE

TM026210

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

Honorable William T. Randall, Chair
Commissioner of Education
State of Colorado
Denver, Colorado

Mary R. Blanton
Attorney
Salisbury, North Carolina

Patsy Cavazos
Principal
W.G. Love Accelerated Elementary School
Houston, Texas

Catherine A. Davidson
Secondary Education Director
Central Kitsap School District
Silverdale, Washington

Edward Donley
Former Chairman
Air Products & Chemicals, Inc.
Allentown, Pennsylvania

Honorable James Edgar
Governor of Illinois
Springfield, Illinois

James E. Ellingson
Fourth-Grade Classroom Teacher
Probstfield Elementary School
Moorhead, Minnesota

Chester E. Finn, Jr.
John M. Olin Fellow
Hudson Institute
Washington, DC

Michael J. Guerra
Executive Director
Secondary Schools Department
National Catholic Educational Association
Washington, DC

Jan B. Loveless
Coordinator of Fund Development
Midland Public Schools
Midland, Michigan

Marilyn McConachie
School Board Member
Glenbrook High Schools
Glenview, Illinois

Jason Millman
Prof. of Educational Research Methodology
Cornell University
Ithaca, New York

Honorable Richard P. Mills
Commissioner of Education
State of New York
Albany, New York

William J. Moloney
Superintendent of Schools
Calvert County Public Schools
Prince Frederick, Maryland

Honorable Annette Morgan
Member
Missouri House of Representatives
Jefferson City, Missouri

Mark D. Musick
President
Southern Regional Education Board
Atlanta, Georgia

Mitsugi Nakashima
President
Hawaii State Board of Education
Honolulu, Hawaii

Michael T. Nettles
Professor of Education & Public Policy
University of Michigan
Ann Arbor, Michigan
and Director
Frederick D. Patterson Research Institute
United Negro College Fund

Honorable Roy Romer
Governor of Colorado
Denver, Colorado

Honorable Edgar D. Ross
Judge
Territorial Court of the Virgin Islands
Christiansted, St. Croix
U.S. Virgin Islands

Fannie L. Simmons
Mathematics Coordinator
District 5 of Lexington/Richland County
Ballentine, South Carolina

Adam Urbanski
President
Rochester Teachers Association
Rochester, New York

Deborah Voltz
Assistant Professor
Department of Special Education
University of Louisville
Louisville, Kentucky

Marilyn A. Whirry
Twelfth-Grade English Teacher
Mira Costa High School
Manhattan Beach, California

Dennie Palmer Wolf
Senior Research Associate
Harvard Graduate School of Education
Cambridge, Massachusetts

Sharon P. Robinson (Ex-Officio)
Assistant Secretary
Office of Educational Research
and Improvement
U.S. Department of Education
Washington, DC

Roy Truby
Executive Director, NAGB
Washington, DC

The NAEP Guide

A Description of the Content and Methods of the 1994 and 1996 Assessments

Nada Ballator

with the invaluable assistance from professional staff at
National Center for Education Statistics,
Educational Testing Service,
National Computer Systems, and Westat, Inc.;
and with gratitude to the authors of the previous edition.

Revised Edition: October 1996

Prepared by EDUCATIONAL TESTING SERVICE under a cooperative agreement with
THE NATIONAL CENTER FOR EDUCATION STATISTICS
Office of Educational Research and Improvement
U.S. Department of Education

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

October 1996

FOR MORE INFORMATION:

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).

The work upon which this publication is based was performed for the National Center for EducationStatistics, Office of Educational Research and Improvement, by Educational Testing Service.

Educational Testing Service is an equal opportunity, affirmative action employer.

Educational Testing Service, ETS, and *the ETS logo* are registered trademarks of Educational Testing Service.

TABLE OF CONTENTS

INTRODUCTION 1

QUESTION 1:
What is NAEP? 3

QUESTION 2:
What subjects does NAEP assess? How are the subjects chosen, and how are the questions for the assessments specified? What subjects were assessed in 1994 and 1996? 8

QUESTION 3:
Who are the students assessed by NAEP? 19

QUESTION 4:
Is participation in the National Assessment of Educational Progress voluntary? Are the data confidential? Are student names and other identifiers available? 23

QUESTION 5:
Can parents examine the questions used by NAEP to assess the achievement of their children in core academic subjects? Why are NAEP questions kept confidential? 26

QUESTION 6:
How does NAEP accommodate students with disabilities and students with limited English proficiency? 28

QUESTION 7:
How many schools and students are involved in the 1996 national and state assessments? When are the data collected during the school year? 31

QUESTION 8:
How are the assessments developed? 35

QUESTION 9:
What innovations in assessment has NAEP developed? 38

QUESTION 10:
What kind of results does NAEP provide? 42

QUESTION 11:
What contextual background data does NAEP provide to help decision makers interpret the achievement results? 46

QUESTION 12:
How does NAEP use matrix sampling to reduce the burden for participating students? What is "focused BIB spiraling" and what are the advantages of using it in NAEP? 50

QUESTION 13:
What are the NAEP procedures for data collection? 53

QUESTION 14:
How does NAEP reliably score and process millions of student-composed responses in an assessment? 56

QUESTION 15:
How does NAEP analyze the assessment results? 63

QUESTION 16:
How does NAEP ensure the comparability of results within the state assessments, and also the comparability between state and national results? 68

QUESTION 17:
What types of reports does NAEP produce? What reports are planned for the 1996 assessment? 70

QUESTION 18:
In what ways can educators apply NAEP resources to their work? How can they use frameworks, questions, and reports? 74

QUESTION 19:
How can NAEP data and assessment results be used for further exploration of education and policy issues? What kinds of technical assistance does NAEP provide? 76

QUESTION 20:
Can NAEP results be linked to international assessment data? 79

QUESTION 21:
Who evaluates and validates NAEP 81

BIBLIOGRAPHY 84

CHART: **Subject Areas Assessed by NAEP** ..... back cover

INTRODUCTION

The National Assessment of Educational Progress (NAEP) is mandated by Congress to survey the educational accomplishments of U.S. students and to monitor changes in those accomplishments. NAEP tracks the educational achievement of fourth-, eighth-, and twelfth-grade students over time in selected content areas. For over 27 years, NAEP has been collecting data with the aim of providing accurate and useful information to educators and policy makers.

How NAEP is organized. The NAEP project employs the full-time equivalent of over 125 people each year, and as many as 5,000 people work on NAEP in one or another capacity. These people work for many different organizations, which must cooperate their efforts to conduct NAEP. Amendments to the authorizing statute for NAEP in 1988 established the present structure for this cooperation. Under the present structure, the Commissioner of Education Statistics, who heads the National Center for Education Statistics (NCES) in the U.S. Department of Education, is responsible for NAEP operations and technical quality control.

The program is governed by the National Assessment Governing Board (NAGB), appointed by the Secretary of Education but independent of the department. The Governing Board, which is authorized to set policy for NAEP, is designed to be broadly representative of NAEP's varied audiences. NAGB selects the subject areas to be assessed and ensures that content is planned through a national consensus process; the Governing Board currently contracts with the Council of Chief State School Officers (CCSSO) for national consensus development. In addition, the Governing Board identifies achievement standards for each subject and grade tested, in conjunction with its contractor, the American College Testing Program (ACT). NAGB also develops guidelines for reporting. NAGB's authority to oversee NAEP and give policy direction to NCES parallels that of the Commissioner of Education Statistics to direct and execute the program.

NCES procures test development and administration services from cooperating private companies. Since 1983, NCES has implemented the assessment through a cooperative agreement awarded to Educational Testing Service (ETS). Under this agreement, ETS has the direct responsibility for developing the assessment instruments,

scoring the student responses, analyzing the data, and reporting the results. NCES has also awarded a cooperative agreement to Westat, Inc., for the selection of school and student samples, training of assessment administrators, and management of field operations (including assessment administration and data collection activities). The printing and distribution of the assessment materials, as well as the scanning and scoring of students' responses is conducted by National Computer Systems (NCS) under subcontract to ETS.

The results of the NAEP assessment are published by NCES and, through the Public Affairs Office of the U.S. Department of Education, released to the media and the public. NCES strives to present the information obtained in the most accurate and useful manner possible by publishing reports tailored to the general public as well as to specific audiences and by making data available to researchers for secondary analyses.

About the Guide. The purpose of this *NAEP Guide* is to provide interested readers with an overview of the scope of the project and to increase the understanding of the philosophical approach, procedures, analysis, and psychometric underpinnings of the NAEP design. This Guide also acquaints readers with NAEP's informational resources, demonstrates the appropriateness of NAEP's design to its role as indicator of national educational achievement, and describes some of the methods used in the 1994 and 1996 assessments.

The Guide follows a question-and-answer format: The most commonly asked questions about NAEP are followed by succinct answers (in italics), with additional details included for background. It is hoped that the Guide will prove useful to the general public, including state policy makers; to state, district, and school education officials who participate in state or national NAEP; and to researchers who use this as their introduction to NAEP.

QUESTION 1:

What is NAEP?

ANSWER:

Often referred to as the Nation's Report Card, the National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. It is the most comprehensive national measure of students' learning outcomes at critical junctures in their school experience.

Since 1969, assessments have been conducted on a regular timetable in a variety of subjects. By making objective information on student performance available to policy makers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the conditions and progress of education. Only information related to academic achievement is collected under this program, and NAEP guarantees the confidentiality of all data related to the individual students assessed and their families.

FURTHER DETAILS:

Overview of NAEP

Over the years, NAEP has evolved to address the many questions asked by policy makers, to the point that NAEP no longer refers to a single assessment. NAEP is now a collection of national and state-level assessments.

NAEP is designed to meet two central goals:

- to reflect the most current and best educational and assessment practices and
- to measure reliably change over time.

Because the program is committed both to creating forward-looking instruments and also to measuring performance trends over time, NAEP uses a multifaceted strategy whereby different nationally representative samples of students participate either in the assessments based on recently developed frameworks and specifications (*main* NAEP) or in the *long-term trend* NAEP assessments.

Main NAEP assessments, designed to follow the forward-looking curriculum frameworks developed by NAGB, must make use of the latest advances in assessment methodology. Indeed, NAEP has pioneered many of these innovations. Main NAEP assessments are constructed to reflect student achievement in a wide range of academic subjects, which are chronicled on the back cover.

In contrast, measuring change over time requires the precise replication of past procedures. The long-term trend assessments are instruments developed in the mid-1980's that are administered every two years in the same form in which they were originally given. These assessments (in four basic subjects) allow NAEP to measure trends that extend back as far as 1969.

The distinction between the two approaches is, of course, not absolute. For example, the mathematics assessment that was administered in 1996 follows a *short*-term trend line begun in 1990 and continued in 1992.

The two components that use (separate) national samples — main and long-term trend — have assessed various subject areas on an annual basis from 1969 to 1980 and on a biennial basis since 1980. State-level NAEP assessments began in 1990 using the same instruments as administered in the main NAEP. State NAEP offers participating states results that may be compared to those of other participating states as well as linked to the main NAEP national sample.

National NAEP

Main assessments. In addressing the first purpose, national main NAEP assessments have periodically measured students' achievement in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects. The national main NAEP in 1994 covered reading, U.S. history, and geography at grades 4, 8, and 12; there were special studies in reading and history. In 1996, main NAEP assessments were conducted in mathematics and science at grades 4, 8, and 12, as well as special studies of advanced mathematics at grades 8 and 12 and advanced science at grade 12. The assessment instruments used over the years have been intentionally designed to be flexible enough to allow NAEP to adapt to changes in curricular and educational approaches. For example, NAEP assessments have come to rely more on constructed-

response questions and the use of calculators. As the content and nature of the main NAEP instruments evolve to match instructional practices, the ability of the main assessments to measure change over time is greatly reduced. However, in situations where the frameworks underlying the assessment remain stable, the main assessment can adequately measure change over short periods of time. In such cases, the national main assessments also help fulfill NAEP's second main purpose, to measure trends over time. However, in some subject areas, the curriculum may undergo many changes in a short time, so that the framework guiding the assessment must be updated too frequently to permit measurement of short-term change, as in the 1990 and 1996 science assessments.

Long-term trend national NAEP. To allow NAEP to measure change over long periods, certain assessments must remain stable so that changes in student achievement can be examined longitudinally. In its long-term trend program, NAEP accomplishes this by administering identical instruments from one assessment year to the next, measuring students' achievement in four basic subjects (mathematics, science, reading, and writing). Unlike the main assessments, the long-term trend instruments do not evolve based on changes in curricula or in educational practices.

The two assessments that comprise the national NAEP — main and long-term trend — use completely separate samples of students and distinct data collection procedures. The main assessments report results for grade samples (for example, for grades 4, 8, and 12); the long-term trend assessments report results for age samples in mathematics, science, and reading (for 9-, 13-, and 17-year-old students), or in the case of the writing long-term assessment, for grade samples (for grades 4, 8, and 11). Both main and long-term trend assessments use samples of students drawn from public and nonpublic schools over the entire nation. (For a discussion of the samples of students assessed by NAEP, see Question 3.) Student and teacher background questionnaires also vary between main and long-term trend assessments, as do many of the analyses employed to produce results. Finally, the results from these two types of assessments are reported separately.

State NAEP

Until 1990, NAEP was a *national* assessment only, reporting information for the nation as a whole and for geographic regions of the country — Northeast, Southeast, Central, and West. However, Congress passed legislation in 1988 authorizing a voluntary Trial State Assessment (TSA) to address the needs of state-level policy makers for reliable data concerning student achievement in their states. Beginning with the 1990 NAEP, the TSA assessed mathematics at grade 8 in 40 states, territories, and other jurisdictions. In 1992 the TSA assessed mathematics (at grades 4 and 8) and reading (at grade 4) in 44 jurisdictions, and in 1994 the TSA assessed reading at grade 4 in 44 participating jurisdictions. The first two state-level NAEP assessments (1990 and 1992) included only public school students. Nonpublic schools were included for the first time in the 1994 TSA. In 1996, for the state-level NAEP, the "Trial" label has been removed based on the recommendations of numerous evaluations of the previous TSA programs; however, the legislation still notes that the state assessments are developmental. The state-level NAEP in 1996 assessed mathematics at grades 4 and 8 and science at grade 8, with 47 jurisdictions participating.

The various subjects assessed over the years by national main, long-term trend, and state-level NAEP assessments are presented in the table on the back cover. The subjects assessed as part of the long-term trend assessment have remained constant since 1990.

Background questionnaires

In addition to administering subject-specific cognitive questions, NAEP collects data from students, their teachers, and their principals that relate to students' school backgrounds and educational activities. Students are asked questions about course taking, homework, and home factors related to instruction. Teachers are asked about their professional qualifications and teaching activities, while principals are asked about school-level practices and policies. This information is related to students' performance on the cognitive portions of the assessments to provide the context for a better understanding of student achievement.

What factors are related to higher scores? Who is teaching our students? How do schools vary in terms of courses offered? These are some of the questions NAEP attempts

to answer through data collected from student, teacher, and school questionnaires. Putting NAEP results (that is, average scale scores or achievement level percentages) in the context of responses from the background questionnaires increases the usefulness of the findings. (For a more detailed discussion of the various background questionnaires, see Question 11.)

QUESTION 2:

What subjects does NAEP assess? How are the subjects chosen, and how are the questions for the assessments specified? What subjects were assessed in 1994 and 1996?

ANSWER:

Since its institution in 1969, NAEP has assessed a variety of academic subjects, including mathematics, science, reading, writing, geography, history, civics, social studies, and the arts. NAEP also assesses long-term trends in mathematics, science, reading, and writing achievement. (A chronological listing of the assessments from 1969 to 1996 and those planned for 1997 and 1998 is on the back cover.)

Since 1988, the National Assessment Governing Board (NAGB) has been responsible for selecting the subjects assessed by NAEP. NAGB also oversees the creation of the frameworks that underlie the assessments and the specifications that guide development of the assessment instruments. The framework for each subject area is determined through a complex process of consensus involving teachers, curriculum specialists, school administrators, parents, and the general public.

The NAEP 1994 national assessment included reading, U.S. history, and geography, plus the customary long-term trend assessments; at the state level, reading was assessed in grade 4. The 1996 assessment included mathematics (at grades 4, 8, and 12 nationally, and grades 4 and 8 at the state level) and science (at grades 4, 8, and 12 nationally, and grade 8 at the state level), as well as the long-term trend assessments. A description of the components of the NAEP assessments is presented in Question 1.

FURTHER DETAILS:

The selection of subjects, development of frameworks, and specification of assessment questions

Subject areas for assessment. The legislation authorizing NAEP assigns NAGB the responsibility for determining the subjects assessed by NAEP. The subjects assessed as part of the NAEP 1994 and 1996 assessments are listed in the following table. The grade or age at which they were assessed is also indicated.

NAEP 1994 and 1996 Assessments

NATIONAL NAEP 1994			
Main Assessments	**Grade 4**	**Grade 8**	**Grade 12**
Geography	✓	✓	✓
Reading	✓	✓	✓
U.S. History	✓	✓	✓
Long-Term Trend	**Age 9**	**Age 13**	**Age 17**
Mathematics	✓	✓	✓
Science	✓	✓	✓
Reading	✓	✓	✓
Writing (grades 4, 8, 11)	✓	✓	✓
TRIAL STATE NAEP 1994	**Grade 4**		
Reading	✓		

NATIONAL NAEP 1996			
Main Assessments	**Grade 4**	**Grade 8**	**Grade 12**
Mathematics	✓	✓	✓
Advanced Targeted Study		✓	✓
Science	✓	✓	✓
Advanced Targeted Study			✓
Long-Term Trend	**Age 9**	**Age 13**	**Age 17**
Mathematics	✓	✓	✓
Science	✓	✓	✓
Reading	✓	✓	✓
Writing (grades 4, 8, 11)	✓	✓	✓
STATE NAEP 1996	**Grade 4**	**Grade 8**	
Mathematics	✓	✓	
Science		✓	

Development of frameworks. The legislation authorizing NAEP assigns NAGB the responsibility for specifying the content of the assessments through an organizing framework specific to each subject. The framework is thus the blueprint that guides the development of the assessment instrument.

To construct the frameworks, NAGB follows a general pattern of consensus development that has evolved over time. As set forth by the legislation establishing the national assessment, the process calls for "active participation of teachers, curriculum specialists, subject matter specialists, local school administrators, parents, and members of the general public."

Development of the frameworks for the assessments has involved

- widespread participation and review by educators and state education officials in the particular field of interest;
- review by steering committees with members representing policy makers, practitioners, and citizens at large;
- participation of the supervisors for the particular subject in the education agencies of all potential participants; and
- reviews by scholars in that field, by NCES staff, and by a policy advisory panel.

Public hearings have been part of the process. More details on this process of consensus, which is specific to each subject, are provided by the *Framework* publications for the NAEP 1994 and 1996 assessments.

Designed to guide the development of assessment instruments, the frameworks cannot encompass everything that is taught in all the classrooms of the nation, much less everything that *should* be taught. Nevertheless, the frameworks attempt to capture the range of subject-specific content and thinking skills that students should possess to deal with the complex issues they will encounter both inside and outside their classrooms. The consensual nature of the process helps ensure that the frameworks will be appropriate for current educational requirements.

The mandate to monitor change requires that the assessments remain flexible to keep pace with changes in educational objectives and curricula: the frameworks must therefore be forward-looking and responsive, balancing current teaching practices with research findings. Evidence of this flexibility is seen in the evolving nature of NAEP assessment instruments, which have changed significantly over the years. Today's instruments, for example, dedicate over half of the testing time to constructed-response questions (questions that require students to compose an answer rather than select one from a list of several possibilities), as opposed to traditional multiple-choice instruments.

Questions and tasks in the assessment instrument. The questions and tasks are developed by subject-matter specialists and testing experts under the direction of Educational Testing Service (ETS) and guided by information contained in the subject-specific frameworks.

For each subject-area assessment, a committee of national experts provides guidance and reviews the questions to ensure that they are appropriate to the framework specifications. Finally, questions that will be part of the NAEP state component are reviewed by the state curriculum and testing directors who make up the NAEP NETWORK. (For more detail on the test development process, see Question 8.)

The subjects assessed in main NAEP in 1994 and 1996

The framework for the NAEP 1994 geography assessment. The geography framework establishes a context for investigating students' knowledge of key aspects of the subject. It is designed to capture the substance and practical applications of geography in an assessment covering both global and U.S. topics. The framework identifies two primary dimensions for specifying questions: a *content* dimension and a *cognitive* dimension.

The content dimension of the framework specifies three general topics for the questions:

Space and Place — knowledge of geography related to particular places on Earth, to spatial patterns on Earth's surface, and to physical and human processes that shape such patterns

Environment and Society — knowledge of geography related to the interactions between environment and society

Spatial Dynamics and Connections — knowledge of geography as it relates to spatial connections among people, places, and regions

In addition to measuring the content, these questions also measured one of three cognitive dimensions:

Knowing
Understanding
Applying

The makeup of the assessment in terms of the three content and three cognitive areas to be measured is described in detail in *Geography Framework for the 1994 National Assessment of Educational Progress*. More information about the framework and the assessment results can be found in *NAEP 1994 Geography: A First Look* as well as in *NAEP 1994 Geography Report Card.*

The framework for the NAEP 1994 reading assessment (both at the national level and at the state level for grade 4). The NAEP 1992 and 1994 reading assessment framework, grounded in current theory, treats reading as a dynamic and complex interaction between and among the reader, the text, and the context of the reading experience. The framework specifies that the assessment address three general types of text and reading situations:

Reading for literary experience — reading novels, short stories, poems, plays, and essays to learn how authors present experiences, interplays among events, emotions, and possibilities

Reading to gain information — reading from newspapers, magazine articles, textbooks, encyclopedias, and catalogues to learn how to acquire information from each specific type of prose

Reading to perform a task — reading documents such as bus schedules, directions for a game, laboratory procedures, or maps, looking for specific information, understanding the information, and finally applying it. (Not assessed at grade 4.)

The framework also delineated, as a second cognitive dimension, four ways in which readers interact with the text to gain meaning from it. These four modes of interaction, called "reading stances," are

Initial Understanding
Developing an Interpretation
Personal Reflection and Response
Critical Stance

The assessment was composed of questions illustrative of one of the purposes for reading as well as one of the modes of interaction.

For further discussion of the framework, see *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress*; for explanation and results, see *NAEP 1994 Reading: A First Look* and *NAEP 1994 Reading Report Card for the Nation and the States.*

The framework for the NAEP 1994 U.S. history assessment. The framework for the U.S. history assessment identified the main ideas and unifying themes of U.S. history. The framework focuses on preparing an examination designed to indicate mastery of the subject matter, which is organized into four central themes to give perspective and meaning to the people, ideas, and events that shaped the nation. These four major themes are as follows:

Change and Continuity in American Democracy: Ideas, Institutions, Practices, and Controversies — the development of American political democracy from colonial times to the present, including basic principles and core civic ideas developed through the American Revolution, the U.S. Constitution, the Civil War, and the struggles over civil rights

The Gathering and Interactions of Peoples, Cultures, and Ideas — the contributions made to the American heritage and the development of American society by people of other cultures

Economic and Technological Changes and Their Relation to Society, Ideas, and Environment — the transformation of the American economy from rural frontier to industrial superpower and its impact on society, the environment, and ideas

The Changing Role of the U.S. in the World — the impact of geography, resources, interests, and ideals on American foreign policy during the movement from isolation to worldwide responsibility

The framework uses a second organizing concept: U.S. history is divided into eight chronological periods. The four major themes set out above are then mirrored in each of the periods.

There is a third organizing concept in the framework: it examines ways of knowing and thinking about U.S. history and then divides these into two general cognitive domains that are used as a guide in exercise development:

Historical knowledge and perspective
Historical analysis and interpretation

For more explanation of this multidimensional framework, see *U.S. History Framework for the 1994 National Assessment of Educational Progress;* for explanation and results, see *NAEP 1994 U.S. History: A First Look* and *NAEP 1994 U.S. History Report Card.*

The framework for the NAEP 1996 mathematics assessment (both at the national level and for grades 4 and 8 at the state level). The mathematics framework is anchored in five broad *strands* of mathematical content that reflect the current National Council of Teachers of Mathematics (NCTM) standards, just as it was for the assessments in 1990 and 1992. NAGB developed the framework to include the NCTM content strands, and they will be included in the scales for reporting the 1996 mathematics results.

The content strands to be measured are

Number Sense, Properties, and Operations — understanding of numbers, operations, and estimation and their application to real-world situations

Measurement — understanding the process of measurement and using numbers and measures to describe and compare mathematical as well as real-world objects

Geometry and Spatial Sense — understanding spatial relationships, including the extension of proportional thinking to similar figures and indirect measurement

Data Analysis, Statistics, and Probability — the skills of collecting, organizing, representing, reading, and interpreting data, as well as statistical concepts to extend these basic skills

> ***Algebra and Functions*** — extending from work with simple patterns at grade 4, to basic algebra concepts at grade 8, to sophisticated analysis at grade 12, and involving not only algebra but also precalculus and some topics from discrete mathematics

These strand divisions are specified to describe the full spectrum of mathematical content to be measured; there is no intention to separate the questions in the mathematics assessment into discrete categories without connections among them.

The questions in the assessment are cross-classified along a second dimension according to which mathematical *ability* they assess:

> ***Conceptual Understanding***
> ***Procedural Knowledge***
> ***Problem Solving***

A third dimension, *power*, has played an increasingly important role in measuring student achievement since the 1990 assessment. The conception of mathematical power refers to the application of the three mathematical abilities listed above within three broader contexts:

> ***Reasoning***
> ***Connections across the content strands***
> ***Communication of meaningful responses to tasks***

The special theme blocks in the NAEP 1996 mathematics assessment are intended to probe this third dimension of mathematical power.

For a detailed discussion of all aspects of the framework, see *Mathematics Framework for the 1996 National Assessment of Educational Progress.* (For descriptions of the reports of the results, see Question 17.)

The framework for the NAEP 1996 science assessment (both at the national level and for grade 8 at the state level). In 1996 NAEP administered a newly developed assessment in science, based on a new framework. This framework can be thought of as a two-dimensional matrix in which the first dimension specifies three science content areas:

> ***Earth Science*** — basic knowledge of the Earth (lithosphere) and various ways to depict its structure, features, and position in space; the hydrosphere, the movement of water, and the water cycle; and the composition of the atmosphere, energy transfer, and weather
>
> ***Physical Science*** — basic knowledge and understanding of the structure of the universe as well as the physical principles operating within it: matter and its transformations, energy and its transformations, and the motion of things
>
> ***Life Science*** — basic understanding of the nature and function of living things within the major concepts of change and evolution: cells and their functions; organisms and their development, cycles, functions and interactions; and the interdependence of life, or ecology

The second dimension, Ways of Knowing and Doing Science, encompasses three subdimensions:

> ***Conceptual Understanding***
> ***Scientific Investigation***
> ***Practical Reasoning***

In addition to the two major dimensions, the framework requires some interdisciplinary exercises that integrate the three fields of science rather than represent separate content. The two categories are:

> ***The Nature of Science***
> ***Themes in Science (models, systems, and patterns of change)***

A significant portion of the 1996 assessment used hands-on performance tasks and in-depth problems to determine students' science knowledge and understanding. For more detail on the science framework, see *Science Framework for the 1996 National*

Assessment of Educational Progress. (For descriptions of the reports of the results, see Question 17.)

Long-term trend subjects assessed in 1994 and 1996

In 1994 and 1996, the long-term trend assessments were identical to the assessments begun in 1984 for reading and writing, and in 1986 for mathematics and science. Note that trend is measured several years before that because of statistical links with previous years' assessments; the back cover shows specific subjects and years in trend measurement.

The framework for the reading long-term trend assessment. This framework is described in *Reading Objectives, 1983-84 Assessment.* There are four objectives for student achievement:

Comprehending What is Read
Extending Comprehension
Managing the Reading Experience
Valuing Reading

The framework for the writing long-term trend assessment. This framework is described in *Writing Objectives: 1984 Assessment* and *Writing Objectives: 1988 Assessment.* There are four objectives for student achievement:

Writing to Accomplish a Variety of Purposes
Managing the Writing Process
Controlling the forms of Written Language
Valuing Writing and Written Works

The framework for the mathematics long-term trend assessment. There are seven content areas, which are described in *Math Objectives, 1985-86 Assessment*:

Fundamental Methods of Mathematics
Discrete Mathematics
Data Organization and Interpretation
Measurement
Geometry
Relations, Functions, and Algebraic Expressions
Numbers and Operations

The framework for the science long-term trend assessment. The science content areas are described in *Science Objectives: 1985-1986 Assessment* and *Science Objectives: 1990 Assessment.* They are:

Life Sciences
Physics
Chemistry
Earth and Space Sciences
History of Science
Nature of Science

The design of the assessments, sampling, and data collection are described in the Procedural Appendix in *NAEP 1994 Trends in Academic Progress.*

QUESTION 3:

Who are the students assessed by NAEP?

ANSWER:

NAEP is a survey assessment that reports on student performance in selected academic areas for specific populations of in-school students or for subgroups of these populations. NAEP results are based on samples of students selected from the populations of interest. NAEP does not, and is not designed to, report on the performance of individual students.

NAEP national and state assessments are given every two years to representative samples of students enrolled in certain grades or at certain ages in public as well as nonpublic schools in the United States. For the main NAEP assessment at the national level, students are selected from grades 4, 8, and 12 in both public and nonpublic schools; state-level NAEP samples were selected from grades 4 and 8, and have included both public and nonpublic school students since the 1994 NAEP assessment.

For the national long-term trend assessments, students are sampled at ages 9, 13, and 17 for science, mathematics, and reading, and for long-term trend assessment of writing, at grades 4, 8, and 11. (For a more complete description of the various components of the assessments, see Question 1.)

FURTHER DETAILS:

Sampling the students in the nation and in the states

Sampling. NAEP has developed complex sampling designs in its attempt to accomplish two sometimes competing goals:

- to produce precise estimates of student performance and
- to maximize the information available given scarce resources, including students' and teachers' time.

Ensuring representativeness of the samples of students assessed

As the "Nation's Report Card," NAEP must report accurate results for populations and subgroups of students (e.g., minority students or students attending nonpublic schools). For accurate results, the relatively small samples of students selected for the NAEP assessments must be truly representative.

The students assessed by national NAEP must be representative of all in-school students in the nation at the three age or grade levels. Similarly, for state-level NAEP, the students assessed in each participating state or other jurisdiction must be representative of in-school populations within their respective jurisdictions. The student samples for state NAEP and for main NAEP (as well as for long-term trend) are different: no student is in more than one assessment. For both national and state samples, the students are selected randomly according to a complex design constructed to reflect the demographic composition of the nation or the states.

Stratification. The samples are chosen from schools selected by using a multistage design involving stratification (that is, classification into groups having similar characteristics). To endure representativeness, the samples must be randomly selected from groups of schools stratified by variables such as region, urbanization, percent minority enrollment, and median household income.

For the national main and long-term trend NAEP assessments, the sampling design has the following stages:

- selection of primary sampling units (PSUs) — these are geographic areas defined as counties or groups of counties,
- selection of schools (both public and nonpublic) within the selected areas, and
- selection of students within schools.

The stratification that is crucial to the representativeness of the sample is begun by selecting PSUs according to region of the country (Northeast, Southeast, Central, and West), and within each region, designating whether the PSU is urban or rural. The PSUs in some regions are classified according to whether their populations contain a high percentage of enrolled Black or Hispanic students. Further strata may be defined, based on median household income, education level of residents over 25 years, or other

demographic characteristics. Once the stratification is accomplished, the schools within a PSU are assigned a probability of selection that is proportional to the number of students per grade in the school.

The sample design for the state NAEP assessments consists of two stages:

- selection of schools (both public and nonpublic) within the selected areas and
- selection of students within schools.

For the state sample, representativeness of the sample is based on classification or stratification of the public schools by urbanization, minority enrollment, and median household income. The nonpublic schools are stratified by type of control (parochial, private, etc.), urban status, and enrollment per grade.

The samples drawn for the state level are completely separate from those drawn for the national level components. Given the voluntary nature of the state-level NAEP assessments, some states may opt not to participate. Therefore, the aggregate of the state samples may not be a representative sample for the nation.

Oversampling. To address the concern that the results for the major subgroups of the populations are accurately reported, *oversampling* is sometimes required. Oversampling refers to sampling particular types of schools at a higher rate than they appear in the population. For example, national NAEP oversamples schools with high minority populations and also nonpublic schools. This helps ensure that the numbers of students from particular racial/ethnic subgroups and students attending nonpublic schools in the samples are adequate for the accurate estimation of subgroup performance. However, for the samples to be representative of the population as a whole, the data provided by students in oversampled schools need to be properly *weighted* when the analyses are carried out. Weighting accounts for the disproportionate representation of certain subgroups due to the oversampling of nonpublic schools and schools with high minority enrollment. By the same token, the lower sampling rates that sometimes occur for very small schools must also be accounted for with weighting. When properly weighted, NAEP data provide results that are representative of performance for the nation as a whole as well as for the subpopulation of interest, despite the disproportional representation of certain types of schools.

A detailed, technical discussion of the sampling procedures used for the NAEP 1994 assessments can be found in Chapter 3 of the *NAEP 1994 Technical Report* (for the national assessments) and also in Chapter 3 of the *Technical Report of the NAEP 1994 Trial State Assessment in Reading* (for the state-level assessment).

QUESTION 4:

Is participation in the National Assessment of Educational Progress voluntary? Are the data confidential? Are student names and other identifiers available?

ANSWER:

By federal law, NAEP is a voluntary assessment. Federal law specifies that the assessment is voluntary for every pupil, every school and school district, and every state. Although a few state legislatures mandate state participation in NAEP, other states choose not to participate. Some school districts, some schools, and some students choose not to participate, with no adverse consequences from the federal government.

The federal government is responsible for charting the academic performance of fourth-, eighth-, and twelfth-grade students as groups, not as individuals. NAEP is not a test of individual achievement as is the Scholastic Achievement Test (SAT) or a state end-of-grade test, but a report card for the population of all students based on a sample from that population.

Federal laws also dictate that NAEP data be kept confidential. After the National Center for Education Statistics (NCES) publishes NAEP reports, NCES makes the data available to researchers, but the data made available to researchers do not include student names or other identifiers. Some of the data, however, might permit clever and resourceful researchers to deduce the identity of some NAEP schools. To obtain access to NAEP data, researchers must swear to keep identities confidential, under penalty of fines and jail terms.

FURTHER DETAILS:

NAEP is a voluntary assessment.

Participation in NAEP is voluntary for states, schools districts, schools, teachers, and students. Parents of students who are selected to participate are given an opportunity to determine whether their child is to be assessed, though local schools determine the procedures for obtaining parental consent.

Participation in NAEP consists of responding both to test questions focused on a particular subject (reading, mathematics, science, etc.) and to background questions about the subject area, classroom practices, school characteristics, and student demographics. Answering either type of question is voluntary. Since NAEP background questions are created to provide useful information for educators and policy makers, student nonresponse reduces the amount of potentially helpful information that can be reported. A wealth of information about the educational environment comes from the many NAEP background questions students and teachers answer.

NAEP is a confidential assessment.

The legislation authorizing NAEP, the National Education Statistics Act of 1994, Title IV of Improving America's Schools Act of 1994, U.S.C. 9010, stipulates in Section 411(c)(2)(A):

> "The Commissioner shall ensure that all personally identifiable information about students, their education performance, and their families, and that information with respect to individual schools, remains confidential, in accordance with Section 552a of title 5, United States Code."

All government and contractor employees who work with NAEP data are sworn to uphold confidentiality laws and are subject to criminal penalties if they fail to do so. Persons who violate the confidentiality law by disclosing the identities of any NAEP respondents are subject to penalties including fines and prison terms.

School staff use the names of students during test administration to assign specific test booklets to the students who were selected to take that assessment. NAEP then assigns a number to the booklet in order to link it to teacher and school data. Once the booklets have been filled out and any absent students are given a makeup test, NAEP has no further administrative use for the student names. NAEP administrators use a tear-off form that breaks the link between the names and ID numbers, leaving the names behind in the school when the booklets are sent to NAEP for scoring. The portion of the form containing student names remains in the school in a secure storage envelope for a few weeks after the assessment, in case the linkage via ID numbers needs to be double-checked. School officials are asked to confirm that the information in the storage envelope has been destroyed by returning a Destruction Notice to NAEP.

NAEP data are available to researchers.

Since public funds are used to collect NAEP data, the data are made available to the public once collected. However, the data made available do not include any names and addresses or other personally identifiable information. In some cases it is possible, though unlikely, that some schools or teachers could be identified, because the data they reported were unusual or unique and could therefore be linked to other data sources that do contain school names. To minimize this risk, NCES suppresses some of the data, and when the risk of disclosure cannot be reduced sufficiently, NCES requires sworn statements by researchers before they are given access to the data that they will not disclose any identifiable information. Researchers violating the confidentiality laws by disclosing the identities of any NAEP respondents are subject to the same penalties — fines and prison terms — as government and contractor employees.

Throughout its 27-year history, NAEP has respected the privacy of individuals and organizations. This trust has never been breached.

QUESTION 5:

Can parents examine the questions used by NAEP to assess the achievement of their children in core academic subjects? Why are NAEP questions kept confidential?

ANSWER:

It is the right of every parent to have access to the educational and measurement materials that their children encounter. Parents who are interested in obtaining information about the assessment questions may review the demonstration booklet that NAEP provides. Under certain conditions arranged in advance, the booklets to be used in the actual assessment can be reviewed by small groups of parents. Arrangements for this can be made with the school principal or the NAEP field supervisor or school coordinator who will make sure that test security is protected.

Most of the questions used in NAEP assessments are kept secure or confidential for the same reasons that the integrity of other tests or assessments used in the school must be protected. This is because questions that measure student achievement over a period of time need to be administered to students who have never seen them before. NAEP does, however, release nearly one-third of the questions used in each assessment, and these are available for public use. All background questions asked of students are readily available for review in the demonstration booklets mentioned above.

FURTHER INFORMATION:

Parental access to NAEP booklets

Because NAEP understands that parents are interested in their children's experiences in school, NAEP provides a demonstration booklet to the school before the assessment is scheduled. This demonstration booklet, which may be reproduced, contains all student background questions plus sample questions. Parents can obtain copies of these demonstration booklets from the school for examination.

Within the limits of staff and resources, there are procedures that enable school administrators and parents to review the NAEP booklets slated for the current assessment. These arrangements must be made in advance of the local administration dates, so that sufficient materials can be available and interested persons can be notified about the time and location for the review. Upon request, NAEP staff will meet with small groups of parents to review the booklets, with the understanding that no NAEP questions will be duplicated, copied, or removed. These requests may be made to the NAEP data collection staff or by contacting the National Center for Education Statistics (NCES) at 202-219-1761. Individuals without children in the assessment who wish to examine secure NAEP questions may contact the U.S. Department of Education's Freedom of Information Act officer at 202-708-4753.

The importance of security to NAEP

In tests that measure student achievement and permit the comparison of this year's scores to scores of students in previous years (for instance, many state end-of-grade tests, the SAT, and NAEP), some questions will be reused for reasons of continuity and for statistical purposes. These questions must be kept secure if they are to assess trends in academic performance accurately and if student performance is to be reported on existing NAEP score scales. Since NAEP is charged with the regular assessment of what the nation's students know and can do, protecting the NAEP assessment from being compromised is essential. A second reason for protecting the questions is more obvious: If students are given prior knowledge of test questions, then schools and parents will not know whether student performance is based on classroom learning or coaching on specific assessment questions.

After every assessment, many questions are released to the public. NAEP reports generally contain samples of actual questions used in the assessment. Released questions can also be used for teaching and research, and they can be obtained from NCES, NAEP Released Exercises, 555 New Jersey Avenue NW, Washington, D.C. 20208-5653. (For other ways in which released NAEP questions are used, see Question 18.)

QUESTION 6:

How does NAEP accommodate students with disabilities and students with limited English proficiency?

ANSWER:

Historically, NAEP allowed the exclusion of students whom school officials judged unable to respond meaningfully to the assessment either because the students had limited English proficiency (LEP) or because they had a severe mental or physical disability. Students with disabilities (SD) often had individual education plans (IEPs) requiring specific accommodations for test taking. Since NAEP did not offer these accommodations, this frequently meant that these students were unable to participate in the NAEP assessment. While uniform criteria were imposed on all sampled schools, the ultimate decision for inclusion or exclusion was left to individual school personnel.

For each student selected for NAEP who has been identified by his or her school as having limited English proficiency or as having a disability, whether or not included in the assessment, the SD/LEP student questionnaire is completed by the teacher or staff member who is most familiar with the particular student.

FURTHER DETAILS:

The goal: Increased inclusion

It is the intent of NAEP that all students selected be assessed. However, some students selected might be expected to have difficulty with the assessment as normally administered because of disabilities or limited English proficiency. Prior to 1996, NAEP rules allowed school officials fairly wide discretion over whether or not to include students in NAEP. Beginning with NAEP assessments in 1996, attempts were made to have more of the students who were classified as SD or LEP included in the assessment. School administrators were encouraged even more than in the past to permit inclusion of the student classified as SD or LEP if there were any doubt about the need to exclude the student.

The specific parameters for participation in the NAEP assessment by students with disabilities and students with limited English proficiency have changed somewhat over time. In the 1995 field test in preparation for the NAEP assessment in 1996, new criteria for classification of students as SD or LEP were implemented. Also in the 1995 field test, additional accommodations were permitted to reduce the rate of exclusion for these students. The National Center for Education Statistics (NCES) strongly supports wider inclusion and is committed to its implementation; the new inclusion policies were tested formally with the NAEP assessment in 1996.

Accommodations to include more students

Accommodations field-tested in 1995. Several strategies for including more students with disabilities in the NAEP assessment were tried in the 1995 field test:

- provision of large-print test booklets and large-face "talking calculators"
- provision of Braille booklets
- accommodations in administration procedures (small-group or one-on-one settings, extended timing, different response modes, etc.) that were based on the specifications for testing in the student's IEP

Two strategies for including more Spanish-speaking LEP students in NAEP assessments were evaluated in the 1995 field test of mathematics:

- Spanish-English bilingual assessments
- Spanish-only assessments

The Spanish-language study was conducted at grade 4 and grade 8. Several blocks of questions were specially selected for translation so that there was good representation of the mathematics content areas and question types, including those that required the use of calculators, rulers, manipulatives, and geometric shapes.

Two booklets at each grade level were produced in a bilingual format so that facing pages were printed with the Spanish version of questions on the left and the English version of the same questions on the right. A third booklet at each grade level was produced in a Spanish-only format. These booklets were administered to students whose native or first language was Spanish. Students were given 20 minutes for each block of questions instead of the usual 15 minutes.

Accommodations implemented in the main NAEP for 1996. The accommodations provided in the field test were implemented in main NAEP for students who would have been excluded in the past. Spanish-speaking students classified as LEP were given the option of using a bilingual test booklet in mathematics in a portion of the sample. In addition, English-Spanish glossaries were provided at all three grades for designated science books. Other accommodations (such as earphones for the hearing impaired, signers for the deaf, magnifying equipment, and translators) were allowed if provided by the school and specified in the student's IEP. A study to determine the impact of the revised inclusion rules for students classified as SD or LEP is taking place with the 1996 assessment.

Students who cannot be included

Students classified as SD or LEP who cannot be accommodated, either by the administrators of NAEP or by the schools, are excluded from the assessment. For each of these excluded students, the appropriate teacher or staff member completes an SD/LEP student questionnaire. As mentioned above, the SD/LEP questionnaire is completed also for students so classified who are included in the assessment. These questionnaires are useful in providing information about differential exclusion rates across disability conditions and across the states. (More detail on the SD/LEP questionnaire is provided in Question 11.)

NAEP has traditionally included more than 90 percent of the students drawn in the sample. Even though the percent of exclusion is now relatively small, NAEP continues to explore ways to reduce the exclusion rate even more, while ensuring the representativeness and generalizability of NAEP results.

QUESTION 7:

How many schools and students are involved in the 1996 national and state assessments? When are the data collected during the school year?

ANSWER:

For the NAEP 1996 assessment, approximately 12,000 schools provided nearly 480,000 students for all components of the national and state assessments. The breakdown by component may be seen in the first of the two tables on the following pages.

The data for the national main NAEP and state NAEP are collected in overlapping times in the winter. The data for the national long-term trend assessments are collected in the fall for age 13 or grade 8, in the winter for age 9 or grade 4, and in the spring for age 17 or grade 11.

FURTHER DETAILS:

Sample selection at different levels

For each assessment year, NAEP selects national samples representing the entire population of U.S. students in grades 4, 8, and 12 for the main assessment, and (for the long-term trend assessment) students at the appropriate ages for those grades. The selection process differs somewhat according to whether the sample of students is intended for the national main assessment, the national long-term trend assessment, or the state-level assessment. It is worth noting that booklets used in the state NAEP are the same as those used in the national main NAEP assessment; the national long-term trend booklets are almost entirely different from the national main and state booklets.

Main and state NAEP. For a given assessment year, all subject areas are not necessarily assessed at all levels. For instance, for the national main NAEP in 1996, students were assessed in both mathematics and science at all three grades. However, for the state-level NAEP in 1996, states could volunteer to have their students in both grade 4 and grade 8 tested in mathematics and their students in grade 8 tested in science. No students at grade 12 were tested at the state level in either subject.

The main assessment required fewer than 100,000 students at all three grade levels and for both subjects, while the state-level assessment, even with only two grade levels being tested, required a total sample of around 350,000 (given the number of states that wished to participate in 1996). Approximately 1,800 schools took part in the 1996 main assessment, while about 10,000 schools contributed their students' time and other resources to the state NAEP.

Long-term trend NAEP. Another component of NAEP is the long-term trend assessments, which are given every two years in the same four subjects to relatively small national samples. Samples of students are drawn for the long-term trend assessments in mathematics and science by their ages (9, 13, and 17 years), and in reading and writing both by grades (grades 4, 8, and 11) and also by age (9, 13, and 17 years). Due to precedent, long-term trend writing results are reported by grade while mathematics, science, and reading long-term trend results are reported by age.

The students selected for the assessment in mathematics and science long-term trend and for the assessment in reading and writing long-term trend are in nonoverlapping samples, as are the approximately 700 schools. This is important because any student sampled can take only one of the NAEP components and because it would be disruptive in a single school to administer several NAEP components throughout the school year.

Schools and students assessed. The table on the following page permits comparison of the numbers of schools and the numbers of students needed for the three major components of NAEP. Note that these three components use separate, non-overlapping samples.

This table shows that 18,000 students in the main sample have administered to them the special mathematics theme and estimation blocks in addition to the main mathematics spiraled BIB. (See Question 12 for a discussion of the administration of NAEP.) The targeted advanced studies in mathematics and science required samples of 2,000 students at each grade at the national level. (For a description of the assessment components, including the targeted advanced studies, see Question 2.)

Total Schools[1] and Students[2] Sampled in the NAEP 1996 Assessments

NAEP COMPONENTS	Total Schools	Total Students			
NATIONAL NAEP	2,473	123,940			
Main Assessments	1,791	94,142	**Grade 4**	**Grade 8**	**Grade 12**
Mathematics			10,832	11,521	10,660
Theme Blocks			4,004	4,227	3,860
Estimation Blocks			2,115	2,244	1,883
Advanced Targeted Study				2,365	2,971
Science			11,578	11,971	11,480
Advanced Targeted Study					2,431
Long-Term Trend	682	29,798	**Age 9/Gr. 4**	**Age 13/Gr.**	**Age 17/Gr.**
Math/Science			5,419	5,658	3,539
Reading/Writing			5,019	5,493	4,670
STATE NAEP	9,710	355,851	**Grade 4**	**Grade 8**	
Mathematics			126,287	112,837	
Science			3,924[3]	112,803	
TOTALS for all components	**12,183**	**479,791**	**169,178**	**269,119**	**41,494**

[1] Numbers are for totals by grade; there may be some overlap from schools with more than one grade.
[2] Numbers are for the total samples rather than for the reporting samples; therefore, these sample sizes will be larger than the sample sizes published in reports of NAEP results. The sample sizes include all who participated in the assessment; students excluded from the assessment because of their SD/LEP status are not included in these numbers.
[3] Department of Defense domestic and overseas schools administered the NAEP science assessment at grade 4 (as well as at grade 8).

Assessment schedules. The typical assessment periods for the three types of NAEP assessments are presented in the following table. The assessment schedules remain constant across assessment years to permit a more accurate measurement of change over time, and to help ensure that the results are comparable. The long-term trend assessment is administered at three times during the school year to samples of students in three age or grade groups, while the national main and state NAEP are both administered in the winter to samples in three different grades.

BEST COPY AVAILABLE

NAEP Assessment Schedule

Grade or Age	FALL	WINTER	SPRING
Grade 4 (main, state) **Age 9/Grade 4** (LTT*)		**Main** **State** **Long-Term Trend**	
Grade 8 (main, state) **Age 13/Grade 8** (LTT*)	**Long-Term Trend**	**Main** **State**	
Grade 12 (main) **Age 17/Grade 11** (LTT*)		**Main**	**Long-Term Trend**

* LTT refers to ages/grades sampled for long-term trend assessments.

Most of the jurisdictions participating in state NAEP assessments for 1996 were assessed during the month of February. By arranging for the month-long state assessment period to fall in the middle of the two-and-a-half-month national assessment period, the state and national assessments are made more comparable by eliminating much of the effect that a time difference might produce.

QUESTION 8:

How are the assessments developed?

ANSWER:

In order to meet the nation's growing need for information on what students know and can do, the NAEP assessment instruments must reflect changes in curricula and instruction in diverse subject areas. This is especially challenging, since change in instructional design and objectives might be implemented at any time in the nation's 100,000 schools.

The NAEP assessments must also measure change over time. This is accomplished through the long-term trend assessments that are given in the same school year as the national main and state assessments.

Development of the assessment instruments, from writing of questions to analysis of field-test results to construction of the final instruments, is a complex process, consuming most of the two-year interval between administrations. There are many opportunities, both before and after the field test, for reviews that help identify any areas requiring revision or augmentation in order to achieve the specifications of the framework and the achievement levels.

FURTHER DETAILS:

The test development process

The instrument development process utilized by NAEP to develop the instruments used in the main component of the national assessment and in the state assessment is summarized in the table below. All newly developed assessment questions and exercises go through all the steps in this process. The many reviews are used to help identify any areas in which the assessment questions or instruments need revision or augmentation to be completely consistent with the framework and the achievement levels. Thus, there are many opportunities for expert input into the construction of a test that adheres as closely as possible to the goals established by the National Assessment Governing Board (NAGB) through the process of consensus described in Question 2.

Summary of the NAEP Instrument Development Process

- Questions and exercises are written by ETS specialists in test development and subject-matter experts both at ETS and elsewhere and then classified according to framework specifications.
- Questions and exercises are reviewed for content concerns by test development staff experienced in the subject area and revised accordingly.
- Questions and exercises are banked in the test development system, along with all classification information.
- Test developer assembles blocks of questions and exercises for field tests according to specifications.
- Specialists review the blocks; blocks undergo mandatory sensitivity review; blocks undergo editorial review.
- Instrument Development Committees convene and review questions and blocks and independently confirm classification codes.
- Classification codes are independently verified by outside groups of content and assessment experts.
- For the state assessment program, the NAEP NETWORK reviews all questions, exercises, blocks, and questionnaires that will be included in the assessment.
- Test developer updates test development version of the questions and exercises in response to committee, NETWORK, and content and assessment expert reviews.
- Electronic offload of blocks is sent to ETS test production services for photocomposition and layout with artwork.
- Camera-ready copy of blocks is proofed and reviewed.
- Field-test clearance package is assembled and submitted to NCES for approval by NCES, NAGB, and, in the case of background questions, the Office of Management and Budget (OMB) and the Information Management Compliance Division (IMCD) of the U.S. Department of Education.
- Changes requested by NCES, NAGB, OMB and the IMCD are made to camera-ready and electronic versions of blocks and resubmitted for clearance.
- Clearance number is obtained for the field test.
- Camera-ready blocks are proofed and approved for printing. Final versions of other assessment materials (including audiotape and videotape) are approved for production.
- Field-test booklets and questionnaires are prepared, proofed, and printed.
- Exercises in blocks are uploaded to the central database.

Field tests are administered.

- Field tests are scored and analyzed; questions suitable for the assessment are selected.
- Blocks selected as suitable for the assessment are reviewed by subject-matter specialists.
- Blocks undergo sensitivity review and editorial review.
- Instrument Development Committees convene and review questions and blocks and independently confirm classification codes.
- Electronic offload of blocks is sent to test production services; camera-ready copy is proofed and reviewed.
- Assessment clearance package is assembled and submitted to NCES for approval by NCES, NAGB, and, in the case of background questions, the Office of Management and Budget (OMB) and the Information Management Compliance Division (IMCD) of the U.S. Department of Education.
- Changes requested by NCES, NAGB, OMB and the IMCD are made to camera-ready and electronic versions and resubmitted for clearance.
- Camera-ready blocks are proofed and approved for printing. Final versions of other assessment materials (including audiotape and videotape) are approved for production.
- Assessment booklets and questionnaires are prepared, proofed, and printed.
- Exercises in blocks are uploaded to the central database.

Assessments are administered.

Questions are written by ETS test development staff and external experts, classified according to framework specifications, reviewed by subject-area and measurement specialists, and then revised and reclassified accordingly. The questions are then assembled

BEST COPY AVAILABLE

into blocks (collections of questions administered to students as a timed unit) for field-testing according to framework specifications.

The blocks are reviewed by subject-area and measurement specialists and given a mandatory sensitivity review to ensure that the assessment will reflect a thoughtful and balanced consideration of all groups of people. External reviewers, including state education agency personnel, also review the questions for appropriateness across regions and for students from a variety of backgrounds. In addition, NAGB is responsible for ensuring that all questions selected for use in NAEP are free from racial, cultural, gender, or regional bias. The National Center for Education Statistics (NCES) also reviews all NAEP questions, and the background questions are subjected to further review by the Office of Management and Budget (OMB) and the Information Management Compliance Division of the Department of Education (IMCD). As a final quality control precaution to monitor against bias, the results for each question are checked empirically after the field test. The empirical check for fairness employs differential item functioning (DIF) analyses. DIF analyses identify questions that are differentially difficult for a particular subgroup of students (who are defined by racial/ethnic group membership or by gender) for reasons that seem unrelated to the overall ability of the students. (For a more detailed discussion of DIF procedures, see Chapter 9 of the *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading*. Discussion of DIF, as well as other aspects of item analysis, can also be found in Question 15.)

The Instrument Development Committee reviews questions and blocks and independently confirms the classification codes. This committee meets four times in the development cycle to consider question format, appropriateness, and cognitive processes being measured; to refine scoring rubrics after the field test; and to review field test results.

The NAEP NETWORK, including assessment directors from the 50 states and Guam, Puerto Rico, the Virgin Islands, the District of Columbia, the Department of Defense Education Activity schools, plus representatives from nonpublic schools, convenes to review all exercises, blocks, and questionnaires that will be included in the state assessment program.

QUESTION 9:

What innovations in assessment has NAEP developed?

ANSWER:

As changes take place in curricula and in the way subjects are taught, NAEP frameworks are updated to reflect these changes. Two new frameworks guided assessments in geography and U.S. history in 1994. In 1993 an enhanced version of the 1990 mathematics framework was developed, taking into account significant developments in mathematics education that occurred since 1989-90; special theme blocks were created for mathematics; and in 1996, special assessments were targeted for students of advanced mathematics and science. Also for 1996, a challenging new framework for science guided the assessment, resulting in the addition of hands-on components as well as theme blocks. The inclusion of more students requiring special accommodations is an overall innovation for all the 1996 assessments. An arts assessment with major emphasis on performance is planned for 1997.

FURTHER DETAILS:

Innovations in recent assessments

NAEP in 1994. Each new framework contains assessment objectives that have been updated to reflect changes in curriculum and instruction and often requires innovations in assessment instrumentation, scoring procedures, and analysis methodology. Even when the same framework guides assessments over several years, shifts in curricular or instructional practice may require that new blocks of questions or performance tasks be field-tested and used on subsequent assessments.

The NAEP 1994 geography assessment was the first full-scale assessment of geography achievement for fourth-, eighth-, and twelfth-grade students. It required students to think critically about geographic issues, such as the physical and human factors that shape patterns of settlement and trade. About 60 percent of the students' test time was used in constructing short or extended answers. These answers were in some cases written responses; in other cases, students produced maps, charts, and graphs. A

wide variety of stimulus materials — including maps, charts, atlases, photographs, tables, and text — were used to measure students' ability to interpret and analyze geographic materials.

The NAEP 1994 reading assessment continued the short-term trend begun in 1992 by presenting longer, "naturally occurring" passages and increasing the proportion of constructed-response questions so that they accounted for about 60 percent of the students' actual total response time at grade 4 and around 75 percent at grade 12.

The *NAEP Reader* was a special study conducted as a part of the NAEP reading assessment in 1992 and 1994. The study was designed to examine students' comprehension of texts when they were permitted to choose from several selections, as compared to students who were assigned a text, as is the usual case. Samples of eighth and twelfth graders were given a booklet containing seven short stories. One sample of students was given 50 minutes to select and read one story and to answer 12 constructed-response questions about the story they chose, while another sample had to read a story assigned to them and answer the same number of questions.

The NAEP 1994 U.S. history assessment also required the students to devote about 60 percent of their time to constructing responses to questions involving a wide variety of primary documents, graphs, political cartoons, maps, time lines, and other materials selected to measure students' ability to interpret and analyze historical matter.

NAEP in 1996. Not only can the assessments be updated for currency, but they also can be augmented with special studies. For instance, NAEP 1996 assessments at the national level included special components designed to extend and enhance the information from the national main mathematics and science assessments. These special components were targeted toward students of advanced mathematics and science.

The advanced mathematics targeted assessment was administered nationally to students in grades 8 and 12 who had taken advanced courses at their grade level. At grade 8, the questions in the advanced blocks focused on elementary algebra, while at grade 12, the focus was on precalculus mathematics. These students completed three blocks of questions — two advanced blocks plus one block containing some of the same questions that were administered to students in the national main assessment. This third block is

designed to permit the performance of advanced students to be linked to that of the eighth- and twelfth-grade students in the national samples.

Other special assessments and studies have been added to the national main NAEP. Theme blocks were developed for the main NAEP 1996 mathematics assessment and administered to about 4,000 students at each of the three grades at the national level. Each student in this sample took a special theme block in addition to completing a booklet from the main blocks of questions, or "BIB." (For a discussion of BIB, see Question 12.)

In these special mathematics theme blocks, the questions are united under a common topic, and students at each grade level take one of the two different topics (with one of the theme blocks overlapping grades 8 and 12). The questions range from multiple-choice to extended response, the aim being to measure students' ability to work at many levels within a single context.

Another type of special block involves estimation problems. A new type of estimation block, containing constructed-response questions, was an innovation for 1996. Data from these special theme and estimation blocks extend the scope of the national main NAEP assessment.

In 1996 the mathematics assessment provided for use of calculators, rulers, protractors, and manipulatives in the main BIB.

The NAEP 1996 science assessment was built according to an innovative new framework specifying that students must use higher order thinking skills for multiple-choice questions, constructed-response exercises, and performance tasks. The science framework is based on the view that scientific knowledge should be organized to provide a structure that connects and creates meaning for factual information. Consequently, the new framework concentrates on students' ability to relate basic concepts to each other as well as their ability to discuss and evaluate approaches to science-related problems.

In the NAEP 1996 science assessment, hands-on tasks were a regular part of the main assessment, and a targeted special assessment was added to measure what students in grade 12 who have completed or were in the process of completing four or more science courses know and can do. This special assessment was administered to a sample of advanced science students who completed a block of questions in each of three

subjects: chemistry, physics, and biology. Each student also completed a fourth block containing questions from all three content areas. Since this block contained questions that were also administered to students in the national main assessment, results from the students of advanced science can be linked to performance on the main assessment.

QUESTION 10:

What kind of results does NAEP provide?

ANSWER:

NAEP provides results such as subject-matter achievement, instructional experiences, and school environment about populations of students (e.g., fourth graders) and subgroups of the population (e.g., male students or Hispanic students). NAEP does not provide individual scores for the students assessed.

In order to make the performance of students more easily understood, the subject-matter achievement is linked to two types of scales. The NAEP scale ranges from 0 to 500; the Achievement Level scale aims at greater simplicity by categorizing students' achievement (in reference to established levels of achievement set for each grade) within the ranges Basic, Proficient, and Advanced. A fourth level, below Basic, is also reported for this scale. Both types of scales are developed independently for each subject, so comparisons of rankings on the same scale cannot be made across subjects. However, these scales greatly facilitate comparisons of performance within a subject from one year to the next, or from one group of students to another in the same grade.

FURTHER DETAILS:

NAEP subject-matter achievement and NAEP contextual variables

National-level NAEP is the "Nation's Report Card," and as such examines the collective performance of the nation's students. State-level NAEP provides similar information for participating jurisdictions. NAEP does not, and is not designed to, report on the performance of individual students. (For a discussion of NAEP samples, see Question 3.) NAEP does report, however, on the overall performance of aggregates of students (e.g., the average geography scale score for eighth-grade students, or the percentage of eighth-grade students performing at or above the *Proficient* level in geography). In addition, NAEP reports on major subgroups of the population defined by demographic factors such as race/ethnicity, gender, level of parental education, location

of the students' school (i.e., central city, urban fringe/large town, or rural/small town), and type of school (i.e., public or nonpublic).

Also, students can be grouped on the basis of information provided by the questionnaires completed by the students, teachers, and school administrators. This permits the examination of student performance in the context of various activities related to education. For instance, some of the many contextual variables from the student, teacher, and school questionnaires used to report results for the NAEP 1994 assessments were course taking, homework, use of textbooks or other instructional materials, use of computers, discussion of school work at home, and television viewing. (For a discussion of the contextual variables, see Question 11.)

For all subject areas, a great deal of information about the knowledge and abilities of students was provided by the NAEP 1994 and 1996 assessments. For both the main component of the national assessments and the state-level assessments, the usefulness of this information was maximized by presenting average scores on the NAEP subject scales as well as the percentages of students attaining specific NAEP achievement levels.

Achievement level scales have been developed by NAGB for each subject at each grade level as a measure of the fit between students' *actual* achievement and the achievement *desired* of them, beginning with the 1990 assessment. Thus NAEP results provide information about what students *know* and *can do* (on the NAEP subject scale), and they also indicate the extent to which their achievement meets expectations of what students *should know* and *should be able to do* (as an achievement level).

Results from the long-term trend component of the national assessment are reported in terms of NAEP subject-area scale scores. Achievement levels have not been developed for use with the long-term trend instruments.

NAEP subject scales. For each of the subject areas assessed, student responses to the NAEP assessment are analyzed to determine the percentage of students responding correctly to each multiple-choice question and the percentage of students performing in each of the score categories for constructed-response questions. (For a discussion of characteristics of NAEP assessment instruments, see Question 2.) Item response theory (IRT) methods are used to produce scales that summarize performance on the primary dimensions of the curricular frameworks used to develop the assessment. For example,

the NAEP 1992 and 1994 reading assessments resulted in the production of three scales: reading for literary experience, reading to gain information, and reading to perform a task. Also produced was a reading composite scale, which was a weighted average of the three scales. The composite scale is the principal metric used in reporting NAEP results.

The scales for each NAEP subject are developed independently of each other. Therefore, results cannot be compared across subjects. (Details of the scaling procedures for the NAEP 1994 assessments are presented in the *NAEP 1994 Technical Report.*)

Achievement levels. For each of the subjects assessed in 1994 and 1996, results are also reported using the achievement levels that were authorized by the NAEP legislation. These achievement levels were defined by a broadly representative panel of teachers, education specialists, and members of the general public, and were adopted by the National Assessment Governing Board (NAGB). The achievement levels are based on collective judgments about what students *should know* and *should be able to do* relative to the body of content reflected in the NAEP assessment framework. (For a brief description of the NAEP frameworks for the 1994 and 1996 assessments, see Question 2.)

For each grade level assessed, three achievement levels are defined: *Basic*, *Proficient*, and *Advanced*. The policy definitions of the three achievement levels are presented below; detailed descriptions of the subject-specific achievement levels appear in the *Report Card* for that subject.

Achievement Level Policy Definitions

Basic	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
Proficient	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
Advanced	This level signifies superior performance.

It should be noted that the setting of achievement levels for the national assessment is relatively new and that procedures for establishing the levels are evolving. Although the reviews of the achievement levels have produced conflicting conclusions,

NAGB believes that use of these developmental achievement level results provide a useful and valuable reporting metric for NAEP. For example, reporting results in one of three achievement levels aids understanding of how the percentage of fourth-grade students who scored at or above the *Advanced* level in reading in 1994 compared to that in 1992, or how the percentage of eighth-grade students who scored at or above the *Proficient* level in mathematics in your state compared to that of the nation.

QUESTION 11:

What contextual background data does NAEP provide to help decision makers interpret the achievement results?

ANSWER:

As part of the assessment, NAEP collects information from participating students, teachers, and principals about hundreds of contextual background variables related to student, teacher, and school characteristics as well as to instructional practices and curricula. In developing the questionnaires, NAEP ensures that the questions do not infringe on respondents' privacy, that they are grounded in research, and that the answers can help inform the debate about educational reform. The questionnaires are either in the form of separately timed blocks of questions in the assessment booklets (as in the three sets of student questions) or else they are printed separately (as in the teacher, school characteristics, and SD/LEP questionnaires).

The four general sources for questions providing context for NAEP results are:

- *student questions*
  - *background characteristics*
  - *subject area experience*
  - *motivation on the assessment*
- *teacher questionnaire*
  - *Part I: background and training*
  - *Part II: classroom-by-classroom information*
- *school characteristics and policies questionnaire*
- *SD/LEP questionnaire*

The development of the questionnaires was directed by a framework using a process similar to that followed in constructing the cognitive questions in the assessment. This process included review by external advisory groups, followed by field-testing and further review.

For the main and state NAEP assessments, the student background questions are in a noncognitive block that differs very little by grade within a subject. The student

subject area questions are in a noncognitive block that probably differs somewhat by grade level within a subject. The questions concerning student motivation are in a noncognitive block that is generally the same (except for subject name) over all three grades and across subjects. However, these three types of noncognitive blocks do not differ among the assessment booklets for a given grade and subject as the cognitive blocks do.

The questionnaires for the teachers of the students being assessed differ according to subject area, and may differ according to grade level. The principal or other chief school official completes one school characteristics and policies questionnaire for each grade with students participating in NAEP.

FURTHER DETAILS:

Students being assessed complete questionnaires.

The **student background questions** include questions about such factors as race/ethnicity, school attendance, and academic expectations. These questions also ask the student for information about factors believed to influence academic performance, such as homework habits, language spoken in the home, and quantities of reading materials in the home. In order to document changes that occur over time, many of these questions about contextual factors remain unchanged over assessment years.

The **student subject area questions** follow the background questionnaire. They gather three categories of information: time spent studying the subject, instructional experiences in the subject, and attitudes and perceptions about the subject. These questions are specific to each subject area, and so can probe in some detail the use of specialized resources, such as calculators in mathematics class.

The **student motivation questions** come at the end of the assessment booklet. These five questions ask students how many questions they thought they got right on the NAEP assessment, to describe how difficult they found the assessment, how hard they tried compared to other tests in that subject, how important it was for them to do well on this assessment, and how familiar they were with the assessment format requiring detailed solutions.

Teachers of assessed students complete questionnaires.

To supplement the information on instruction reported by students, the teacher for the subject in which the student is being assessed completes a questionnaire concerning instructional practices, teaching background, and other related information.

The **teacher questionnaire**, **Part I**, on background and general training, has questions concerning race/ethnicity, years of teaching experience, certification, degrees, major and minor fields of study, coursework in education, coursework in specific subject areas, amount of in-service training, extent of control over instructional issues, and availability of resources for the classroom. **Part II** of the teacher questionnaire, on training in the subject area and classroom instructional information, contains about 45 questions concerning the teacher's exposure to various issues related to the subject and to the teaching of the subject. It also asks about such information as pre- and in-service training, the ability level of the students in the class, whether they were assigned to the class by ability level, the length of homework assignments, and use of particular resources.

Principals or other officials in participating schools complete questionnaires.

The **school characteristics and policies questionnaire** is given to the principal or other official of each school that participates in NAEP. This questionnaire has over 60 questions that collect information about the background and characteristics of the schools, including the length of the school day and year, school enrollment, absenteeism, dropout rates, size and composition of teaching staff, policies about tracking, curricula, testing practices, special priorities and school-wide programs and problems. This questionnaire also collects information about availability of resources, policies for parental involvement, special services, and community services.

The **SD/LEP questionnaire** is completed by the teachers of those students who were selected to participate in the assessment and who were classified by the school as students having limited English proficiency (LEP) or who were classified as students with disabilities (SD) who have an individual education plan (IEP) or equivalent classification. For each student classified as SD or LEP and selected for the NAEP assessment, school personnel are required to complete a questionnaire about the background and

characteristics of that student and the reason for the classification. For students classified as SD, the questionnaire requests information about students' functional grade levels, mainstreaming, and special education programs. For students classified as LEP, it asks about students' native language, time spent in special education and language programs, and the level of the students' English language proficiency.

NAEP policy states that if there is any doubt about a student's ability to participate, that student should be *included* in the assessment. Beginning with the NAEP 1996 assessments, more accommodations are being made for both categories of students. (For more information on students classified as SD or LEP and accommodations for them, see Question 6.)

QUESTION 12:

How does NAEP use matrix sampling to reduce the burden for participating students? What is "focused BIB spiraling" and what are the advantages of using it in NAEP?

ANSWER:

The complex frameworks that guide the NAEP assessments typically require several hundreds of questions to test reliably each of the many specifications of the framework. Administering the entire collection of cognitive questions to each student would require far too many hours of student time to be practical.

If the assessment is planned so that there are (for example) 20 to 30 test booklets for a particular subject, each with a different selection of the cognitive blocks, then it becomes possible to assess the entire subject area while requiring a more reasonable amount of testing time for each student. Matrix sampling accomplishes this. In matrix sampling, different portions of the total pool of cognitive questions are put into booklets and administered to different but equivalent samples of students. Thus, not all students are asked to answer all questions, and yet there is complete coverage of the subject being assessed while assessment time per student is minimized.

The NAEP test booklets are composed of background questions and cognitive questions, arranged in blocks. (Blocks are separately timed collections of questions). The blocks of background questions are the same for every student in the same grade for the same subject, but the blocks of cognitive questions (and their sequence in the booklet) vary according to booklet version. This planned variation in test booklets is necessary because NAEP needs to produce precise estimates of student performance while maximizing the information available given scarce resources, including students' and teachers' time.

NAEP uses a variant of matrix sampling called focused-balanced incomplete block (BIB) spiraling for the assessment. Using BIB spiraling, NAEP samples enough students to obtain precise results for each question while requiring (for most subjects) an

average of about an hour and a half of each student's time for the cognitive blocks and also the background questions.

The "focused" part of NAEP's matrix sampling method requires that each student respond to questions from only one subject area being assessed. The "BIB" (balanced incomplete block) part of the method ensures that students receive different interlocking parts of assessment forms, enabling NAEP to check for any aberrant interactions between the different samples of students and the different sets of assessment questions. Spiraling refers to the manner in which the booklets are assigned to pupils, in order to ensure that any group of students is assessed with approximately equal numbers of the different booklet versions.

FURTHER DETAILS:

Balanced incomplete block spiraling

Matrix sampling is often used in test construction when there are more objectives to be assessed than there is time to test them. There are other forms of matrix sampling besides BIB spiraling, but many have drawbacks that make them unsuitable for use in NAEP. For instance, with simple matrix sampling, separate sets of questions are confined to particular booklets; however, having each booklet contain a set of specific questions means that context and order effects must be considered. Simple matrix sampling may occasionally be used in NAEP if it suits the needs of certain assessment questions — but the more sophisticated BIB method is used for the majority of questions to produce data that are more useful because they are relatively free from placement effects.

In a simple matrix design, the same subject-area questions would always be last in the assessment booklet. Student mastery of these questions might be underestimated because of fatigue or overestimated due to practice effects. In the NAEP BIB design, the cognitive blocks are balanced; that is, each cognitive block appears an equal number of times in every position. Each cognitive block also appears paired with every other cognitive block in at least one test booklet.

A simplified example is presented below. This example is based on the NAEP 1990 mathematics design. The full sample of students is divided into seven equivalent groups. Each group of students is then assigned one of the seven booklets. Note that each

cognitive block appears exactly once in each of three possible positions and each block is paired exactly once with every other block. (We show only the cognitive blocks, although test booklets always contain questionnaire blocks as well.) The NAEP BIB design varies according to subject area.

A Model of BIB Matrix Sampling

Booklet version	Position 1 cognitive block	Position 2 cognitive block	Position 3 cognitive block
1	A	B	D
2	B	C	E
3	C	D	F
4	D	E	G
5	E	F	A
6	F	G	B
7	G	A	C

It is worth noting that for the national main and state NAEP 1996 assessments, mathematics followed the BIB matrix (not considering the theme and estimation blocks), while science used a variant of the BIB in that each booklet always had the hands-on task at the end of the cognitive section.

BIB spiraling is used in NAEP even though it requires the printing of a greater variety of test booklets. Each version of the assessment booklet must appear approximately the same number of times in the sample and be administered to equivalent subgroups of the full sample. To accomplish this, the booklets are packed in spiral order (in the above example, one each of booklets 1 through 7, then 1 through 7 again, and so on). These booklets are randomly assigned by the test coordinator to the students in each test administration session. This spiraled distribution of the booklets promotes comparable sample sizes for each version of the booklet, ensures that these samples are randomly equivalent, and also reduces the likelihood that students are seated within viewing distance of an identical booklet.

QUESTION 13:

What are the NAEP procedures for data collection?

ANSWER:

For the NAEP 1996 assessments, data collection operations are specified under a cooperative agreement between Westat, Inc., and the National Center for Education Statistics (NCES) to carry out the sampling and data collection operations for the national NAEP (including long-term trend) and for the state NAEP.

The national assessment is administered by Westat field staff, who go through extensive training. For the state-level NAEP, each participating state is responsible for data collection, while uniformity of procedures across states is achieved through training, supervision, and quality control monitoring by Westat.

The complex process by which NAEP data are collected is monitored closely; the tight control on this process contributes much to the quality — and thus to the comparability — of the national main and state assessments and the results they produce.

FURTHER DETAILS:

Organization and supervision of data collection

NAEP relies heavily on the goodwill of school administrators, and obtaining school cooperation requires substantial time and energy. For any school selected, participation in NAEP is voluntary. The sampled schools are notified of their selection through a series of mailings, including letters to the members of the Council of Chief State School Officers and district superintendents. In addition, informational materials are sent and procedures are explained at in-person introductory meetings.

Westat, Inc., has the following primary responsibilities related to field administration:

- selecting the sample of schools and students
- developing the administration procedures, manuals, and samples
- hiring and training staff to conduct the assessments (for national NAEP)
- training state personnel to conduct assessments (for state-level NAEP)
- conducting an extensive quality assurance program

For the national main and trend NAEP assessments, Westat, Inc., hires and trains approximately 85 field staff to collect the data; the field staff complete all associated paperwork, thus reducing the burden on participating schools.

For the state-level assessment, the NAEP legislation requires that each participating jurisdiction provide for data-collection activities, such as gaining the cooperation of sampled schools and assigning personnel to conduct the assessment according to standardized procedures. Westat employs and trains state supervisors to work with state-appointed coordinators who are responsible for carrying out the organizational tasks related to the state assessment. The individual schools, and the assessment administrators designated by them, are responsible for tasks such as preparing lists of enrolled students for the sampled grade, selecting the students to be assessed, distributing the teacher, school, and SD/LEP questionnaires, and actually administering the assessment. In addition to training the local administrators, Westat provides quality control across states by monitoring 25 percent of sessions in states that have given a NAEP assessment before and 50 percent of the sessions held in states that are new to NAEP. Security of the assessment materials and uniformity of administration are high priorities for NAEP. In the past, the quality control monitors have reported no instances in which there were serious breaches of the procedures or major problems that could jeopardize the validity of the assessment.

After each assessment session, Westat staff interview the assessment administrators by telephone. A final quality control step is to hold a debriefing meeting with the state supervisors to get feedback on the assessment that will help improve procedures, documentation, and training for future assessments.

Management of assessment materials

Under the direction of Educational Testing Service (ETS) staff, National Computer Systems (NCS) produces the many materials required for the NAEP assessments. With few exceptions, the assessment booklets, questionnaires, and tracking forms for main and state NAEP are printed by NCS. NCS prints identifying bar codes with numbers on the booklets and questionnaires, preassigns the booklets to sessions, and then prints the booklet numbers on the administration schedule. This activity improves the accuracy of data collection and is an important part of the spiraled distribution within an assessment session. (For information on matrix sampling using the BIB spiral, see Question 12.)

This preassignment of numbered test booklets to a session and the printing of the booklet numbers on the administration schedule are part of the systematic protection of student confidentiality. Student names are written on small sticky notes which are temporarily applied to the booklets; sticky notes are removed before booklets are given to the students, and then the sticky notes are destroyed. Further, the administration schedules are perforated for easy removal of all student and teacher names; the names are never sent to NAEP. (For more information on confidentiality, see Question 4.)

NCS conducts all receipt control, data preparation and processing, scanning, and scoring activities for the NAEP 1996 assessment. NCS uses an image-processing and scoring system specially designed for NAEP to scan both the multiple-choice selections and the handwritten student responses plus other data provided by students, teachers, and administrators. The introduction of this image-based scoring system during the 1994 assessment virtually eliminated paper handling during scoring. The system also permits on-line monitoring of scoring reliability and the introduction of recalibration methods as discussed below.

QUESTION 14:

How does NAEP reliably score and process millions of student-composed responses in an assessment?

ANSWER:

Scoring around 9 million constructed responses with a high level of reliability, and within an accelerated time frame, was central to the success of the NAEP 1996 assessments. To accomplish this objective, Educational Testing Service (ETS) and National Computer Systems (NCS) performed the following steps:

- *developed focused, explicit scoring guides that match the criteria emphasized in the assessment frameworks*
- *recruited approximately 675 highly qualified and experienced scorers and leaders, trained them, and verified scorer quality through qualifying tests*
- *employed an image processing and scoring system capable of routing the large volume of student responses to the scorers*
- *monitored scorer consistency through ongoing reliability checks; assessed the quality of scorer decision-making through daily backreading*
- *documented all training, scoring, and quality control procedures for the 1996 technical reports*

Both the 1994 and 1996 NAEP assessments contained a mixture of constructed response and multiple-choice questions. The constructed responses were scored through the image processing system, while the responses to the multiple-choice questions were scored by scanning the student responses in the test booklets.

FURTHER DETAILS:

Development of scoring guides

The scoring guides used for the 1996 assessment were developed by ETS staff, following a multistage process.

- First in this process was the articulation of the scoring criteria and the development of an initial version of the scoring guides concurrent with the development of the constructed-response tasks.
- These guides were then subjected, along with the assessment instrument, to a series of internal and external reviews by subject area and measurement specialists, the Instrument Development Committees, the National Center for Educational Statistics (NCES), and the National Assessment Governing Board (NAGB).
- All scoring guides were reviewed to ensure that they included criteria consistent with the wording of the questions. The guides were also reviewed to confirm that they were concise, explicit, and clear and that they reflected the criteria outlined in the assessment framework.
- Next, the guides were used to score the student responses gathered in the field test. The results from this field test were used by the committees and by ETS staff to further refine the guides.
- The last stage in the refinement of the scoring guides took place during the preparation of training materials in the spring of 1996. Assessment specialists from ETS selected examples of students' responses from the actual assessment for each level of performance specified in the guides.

This selection of exemplars provided the final opportunity to refine the wording in the scoring guides, develop additional training materials, and make certain that the guides accurately represented the criteria articulated in the assessment framework.

The exemplars helped make explicit the committees' interpretations of each performance level described in the scoring guides, and also served to illustrate the full range of achievement under consideration. During the scoring of students' answers and

demonstrations, the exemplars helped anchor scorers' interpretations of the scoring guides, thereby ensuring the accurate and reliable scoring of a range of diverse responses.

Recruitment and training of scorers

The recruitment of highly qualified scorers to evaluate students' responses is crucial to the success of the assessment. With this in mind, ETS and NCS employed a five-stage model for selecting and training scorers.

The first stage involved selecting qualified scorers and designing an effective pattern of work distribution. Because scoring involves internalizing complex criteria and applying these criteria to a wide range of student performance, scorers must have knowledge of the subject matter, intellectual maturity, and prior experience with students' work. For the scorers who worked on advanced science and mathematics blocks, a master's degree or experience in the appropriate field or both was preferred. NCS recruited scorers from their pool of experienced raters, as well as through advertisements in area newspapers. Potential scorers took a subject-area examination and engaged in a simulated scoring as well as a series of interviews before the decision to hire was made.

In the second stage, scorers received orientation to the project and training in how to use the image system. This orientation included an in-depth presentation of the goals of NAEP and the framework of the assessment.

The third stage was the preparation of materials to train the scorers, including the selection of examples. To provide scorers with clear examples of each level in the scoring guides, teams of ETS trainers and NCS table leaders read through hundreds of student responses in assessment booklets and photocopied selected responses. Subsequently, multiple copies of the collated sets of papers were made available for training.

In selecting booklets from which the exemplar responses were chosen, the NCS scoring specialists supervising each subject area ensured the equitable representation of the wide range of schools that participated in the assessment. Thus, responses from all race/ethnic and gender groups, all geographic regions, and all types of communities were used during training.

The fourth stage was the actual training of scorers, who were trained by ETS and NCS subject-area specialists using the following procedures:

- presentation and discussion of task to be scored and task rationale; for the hands-on science tasks, this involved having scorers work with the same kits and instructions as had been used by the students
- presentation of scoring guide and the anchor responses
- discussion of the rationale behind the scoring guide, focusing on the criteria that distinguish between the levels of the guide
- practice scoring of a common set of sample student responses
- group discussion of each response contained in the practice scoring
- continuation of the practice steps until the result is a common understanding of how to apply the scoring guide to students' responses

In the last stage, scorers who would be reading the questions requiring long constructed responses worked through a qualification round to ensure that they could reliably score students' responses for extended-response exercises. At every stage, ETS and NCS closely monitored the quality of scorer selection, scorer training, and scorer qualification.

Scoring using the image-based system

The "paper-free" image-based scoring system was designed to accommodate the special needs of NAEP assessments while eliminating many of the complexities found in the paper-based training and scoring approach used through 1992. This image processing system creates electronic pictures of the student responses and projects them onto scorers' computer screens. The new system was used successfully for the NAEP 1994 assessments in history, geography, and reading as well as in the 1995 field tests of science, mathematics, and the arts. In this system, the entire student response booklet is scanned. This includes students' constructed responses as well as their responses to questions in the noncognitive blocks (such as student background questions) and to the multiple-choice questions in the cognitive blocks. The scanned constructed responses are then stored in the system for easy presentation when needed by the scorer. The scorer sees each scanned

student response on a large computer monitor; the selection of possible scores is shown on the same display for quick and accurate scoring.

This system, developed by NCS, greatly facilitated the training and scoring processes for constructed responses because the images of the students' responses could be distributed easily to the appropriate scorers. The system also allowed ETS and NCS staff to monitor scorers' activities in real time. Thus, problems could be identified as they occurred, and solutions could be implemented expeditiously.

The imaging system greatly enhanced scorer reliability for the constructed responses by providing tools for monitoring the accuracy of each scorer. Also, in order to prevent any drift in the scores assigned to certain questions, calibration sets could be systematically collated and presented, giving trainers maximum control and flexibility in preventing drift. This tool was especially useful for questions given at the state level, which involved scoring over 30,000 responses per question. The ability to prevent drift and monitor potential problems, while a team of scorers was evaluating the same question for a period as long as five days, was crucial to maintaining the high quality of scoring.

The image-based scoring system further allowed responses to an exercise that had been spiraled into different test booklets to be grouped, so that all responses to a particular exercise could be scored continuously until finished. In an assessment such as NAEP, which utilizes a balanced incomplete block (BIB) design, working on one question at a time by grouping the responses of all students improves the validity and reliability of scorers' judgments as compared to the previous system of scoring an entire block of several questions at a time.

Rater (scorer) reliability

Rater reliability refers to the consistency with which individual scorers assign a score to a question; this consistency is very important to the success of the NAEP assessments. ETS and NCS employ three methods for monitoring reliability of scoring.

The first method is ongoing *backreading* by team leaders of each scorer's work to confirm that individual scorers are able to apply the scoring criteria reliably across a large number of student responses and that they do so consistently across time. In order to

ensure this, team leaders "backread," or evaluate, approximately 10 percent of each scorer's work.

Also, each group of scorers performs daily *calibration* scoring so the table leader can make sure that there is no drift in the scores assigned. Whenever scorers have had a break of more than 15 minutes (i.e., after lunch break, at the start of the workday, after a fire drill), they scored a calibration set to reacquaint themselves with the scoring criteria as it is applied to the range of student responses.

Lastly, the generation of on-demand *interrater reliability* statistics confirms, on a large scale, the degree of consistency and reliability of each scorer. The need for great consistency of performance among scorers is paramount if the assessment is to produce meaningful results. Accordingly, ETS and NCS have built into the image-based scoring system a means for producing detailed, ongoing interrater reliability reports during the scoring process. These reports, produced on-line and in real time, include information about overall reliability of scoring for individual questions as well as the reliability of individual scorers. They also provide information about the percentage of exact agreement and indicate precisely where disagreements on score distinctions might be occurring. Throughout scoring, ETS coordinators monitor reliability rates for their designated subject areas. This enables them to identify a problem as it is happening, and to retrain individual scorers or groups of scorers as needed.

It should be noted that the measurement of trend in student achievement, whether part of the long-term trend assessment or the short-term trend in the national or state NAEP, involves special concerns for scoring. For either type of trend question, training must make use of the same materials and procedures as in the previous assessment year. This type of scoring involves monitoring reliability rates both within the current assessment year as well as across assessment years. To minimize differences in scoring between assessments, scorers are trained using a selection of papers obtained in the previous assessment. However, experience indicates that, despite consistent score standards and extensive training, there could be some discrepancies in the scoring patterns between different assessments. Thus, a random sample of 20 to 25 percent of the responses from the prior assessment is systematically interspersed among the current responses for rescoring. The results are then used to determine the degree of scoring

agreement between the current and previous assessments. If necessary, current assessment results are then adjusted to account for any differences.

Documenting the process

All aspects of scoring students' answers to constructed-response questions were fully documented. In addition to warehousing the actual student booklets, NCS keeps master files of all training materials and reliability reports. NCS recorded all of the procedures employed to assemble training packets, train scorers, and conduct scoring in the scoring reports. Also included in the scoring reports are all methods used to ensure reader consistency, all reliability data, and all quality control measures. ETS also keeps master files and records the basic scoring procedures and outcomes in the technical report.

QUESTION 15:

How does NAEP analyze the assessment results?

ANSWER:

Before the data are analyzed, the responses from the groups of students assessed are assigned sampling weights to ensure that these groups are represented in NAEP results according to their actual percentage of the school population for the three grades assessed.

The analysis of national and state NAEP data uses these weights and is accomplished in two major phases. In the scaling phase, item response theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question. In the proficiency estimation phase, the results of the scaling are used to produce estimates of achievement levels. The subsequent analyses relate these achievement results to the numerous background variables collected by NAEP.

Because of the importance of the data (both in terms of the amount of money expended to obtain it as well as the reliance placed on the accuracy of the reports based upon it), the scaling and analysis of NAEP data are conducted in a careful manner that includes extensive quality-control checks.

FURTHER DETAILS:

Weighting

Sampling weights are assigned to the responses from groups of students to adjust for oversampling or undersampling from a particular group. For instance, census data on the percent of Hispanic students would be used to assign a weight adjusting the NAEP sample so that it is representative of the nation. That is, the weight assigned a student's responses is the inverse of the probability of that student's selection into the sample. Through weighting, no response is discarded, and each contributes to the results according to the number of students represented by the student assessed. Weighting is also used to adjust for school and student nonresponse (since it can't be assumed that data

are missing randomly), and for other situations. All of the NAEP analyses described below are calculated using these sampling weights.

Scaling and proficiency estimation

NAEP uses IRT methods to produce scales that summarize the results obtained for each content area. Group-level statistics (such as average scores, or percentages of students exceeding specific cut scores) are the principal results reported by NAEP. However, NAEP reports also include a wide range of analyses, many of which examine the relation between these group-level statistics and important demographic, experimental, and instructional variables.

The reporting requirements of the national and state assessments as well as the large number of background variables associated with each assessment require that thousands of analyses be carried out. The procedures NAEP uses to conduct the analyses have been developed because they produce accurate results even with the need to limit the testing burden on students. In addition, these procedures provide data that are readily available for use in secondary analyses.

The following steps are used to generate scale-score data files suitable for analysis:

- Immediately after receipt of the computer files containing students' responses, all cognitive and noncognitive questions are subjected to an extensive item analysis. The item analysis results are reviewed by project staff in search of anomalies that might signal unusual results or errors in creating the database. In parallel with this item analysis, each cognitive question is examined for indications of differential item functioning (DIF).
- After completion of the item analysis and DIF analyses, the IRT scaling phase of the work is begun. Estimates of the parameters of the IRT model are obtained for each question, with separate scales being established for each of several predefined content areas specified in the assessment framework. For instance, for mathematics in 1996 (as in 1990 and 1992) there are five content area scales; for science in 1996, there are three content area scales. (These content areas are described in Question 2.) Because the item parameters

determine the representation of each question in the content area scales, careful checks are made by the psychometric staff to verify that the IRT scaling model provides an acceptable representation of the student responses to the questions. In particular, the fit of the model is examined, by question, for major demographic subgroups as well as for each state in the state-level assessment. Because data collection for the state assessments differs from that for the national assessment, item-parameter estimation is performed separately for state data and national data.

- In the proficiency estimation phase, plausible values of content area scale proficiency scores are generated for each student participating in the assessment. The plausible-values technology, which also uses student background information, allows for *more accurate estimates* of the performance of subpopulations and more appropriate estimates of the variability of those estimates than does the standard (and much simpler) procedure of estimating a standard proficiency score for each student based only on responses to the questions. A series of careful quality-control steps are taken in constructing the plausible values to ensure the accuracy of subpopulation estimates based on these plausible values. The construction of plausible values is conducted separately for each state participating in the state assessment and for each national sample.
- The final step in the analysis involves linking results from the current year to the scales from previous assessments. For national assessments, results are linked to the scales used in previous NAEP assessments of the same subject domain. For state assessments, results in the current year are linked to those for the nation. The selection of the linking function (which is anticipated to be linear) and its adequacy are determined by comparisons of the distributions of scores for the main and state-level assessments.

Plausible values. When procedures and test questions are essentially constant across administrations, as in traditional standardized testing programs, the error structure underlying such a test also remains constant and relatively simple estimation procedures can be used. Because error in the test will be the same in each administration — for

example, error in measuring differences in performance between females and males — any increase or decrease in gender differences can be assumed to be real and not an artifact of the test. NAEP, however, continually changes its test length, test difficulty, and balance of content in order to provide up-to-date, relevant information for policy makers. NAEP needs to use methods that can accommodate substantial updating from assessment to assessment while remaining sensitive enough to measure small but real changes in student performance. The method of plausible values, while complex, meets these needs and is currently the best way to control for the variations in test properties caused by changes in test length, content, and difficulty. Plausible values methodology allows NAEP to produce accurate and statistically unbiased estimates of population characteristics.

The essential idea of plausible values methodology is to represent what the true proficiency for an individual might have been, had it been observed, with a small number of random draws from an empirically derived distribution of proficiency values that is conditional on the observed values of the assessment items and on background variables for each sampled student. The random draws from the distribution can be considered to be representative values from the distribution of potential scale scores for all students in the population with similar characteristics and identical patterns of item responses. The several draws from the distribution are different from one another in a way that quantifies the degree of precision (the width of the spread) in the underlying distribution of possible proficiencies that could have generated the observed performances on the items.

For the full array of NAEP data, the plausible-values approach takes less time than would numerous single runs of a specialized program estimating the proficiency scores of a single subgroup. In the specialized procedure, distributions of performance for a subpopulation can be obtained in narrowly defined content areas without generating plausible values for individual students. However, although NAEP could estimate proficiency distributions without generating plausible values, this would have to be done separately for each subpopulation in each report. Plausible values extend the specialized methodology to handle not just one subpopulation at a time, but all the potential interrelationships among proficiency scales and background variables. The plausible-values approach solves the estimation problem once — albeit with more work than any

one or two or even ten of the simpler single runs — and permits completion of the hundreds of analyses required by the extensive number of NAEP background variables in less time than conducting practically endless separate estimations.

Finally, as a natural by-product, the same student-level data provided by the plausible-values approach, and upon which the NAEP reports are based, can be used by secondary researchers, who can also carry out the full range of NAEP analyses. The specialized approach, in contrast, does not yield detailed student-level information for use by a broad range of secondary researchers relying on standard statistical packages.

In summary, the NAEP implementation of IRT analysis provides, in an efficient way, for extensive, detailed analyses by NAEP staff and by secondary analysts of data that are not biased and have been subjected to numerous quality control steps. The plausible-values scaling technology is at the heart of NAEP's ability to perform the second phase analyses and report the type of results contained in, for example, *NAEP 1994 Trends in Academic Progress.*

QUESTION 16:

How does NAEP ensure the comparability of results within the state assessments and also the comparability between state and national results?

ANSWER:

The closely monitored and standardized process by which NAEP data are collected is described in Question 13. This tight control over the data collection process contributes much to the comparability of the results produced by the main and the state assessments.

The national main NAEP uses the same assessment booklets as the state NAEP. The national and state NAEP assessments are administered during overlapping times (see Question 7), but the administration procedures differ somewhat. To further ensure comparability, statistical equating procedures are used to link the results from both main and state NAEP components to a common scale. The degree of accuracy of this process is confirmed by comparing the distributions of student ability in both samples, thereby justifying reporting the results from the national and state components on the same scale.

FURTHER DETAILS:

Equating national and state assessments

A major purpose of state-level NAEP is to give each participating jurisdiction the ability to compare its results with those for the region of the country in which that jurisdiction is located, as well as with the results for the nation as a whole. Initially, however, data from the state- and national-level assessments must be scaled separately for two major reasons:

- a difference in administration procedures (Westat staff collect the data for the national-level assessment, while data collection for the state-level assessment are the responsibility of individual jurisdictions) and
- potential motivational differences between the samples of students participating in the national and state assessments.

For meaningful comparisons, the national and state assessments need to be equated so they can be reported on a common scale. A major aspect of the equating involves those parts of national and state samples that represent a population common to both components. For instance, in 1996, mathematics was assessed at the fourth- and eighth-grade levels nationally as well as state-by-state. Therefore, two independent samples from this population exist in NAEP:

- the students tested in the national assessment that come from the states that participated in state NAEP (referred to as the state comparison sample or SCS)
- the aggregation of the state-level samples obtained in the state NAEP (referred to as the state aggregate sample or SAS).

Thus, equating and scaling of the national-level with state-level results for the NAEP 1996 mathematics assessment would be accomplished through the common populations of fourth- and eighth-grade public school students. Separate scales from the two components are developed and subsequently linked by setting scale score means and standard deviations equal for the SCS and SAS samples.

Verifying comparability

Following linking, analyses are carried out to verify the degree to which the process produced comparable scales for both state and national results. One way to determine such comparability is to compare the shapes of the distributions for the SCS and SAS samples. To examine this, each scale is divided into 10-point intervals, and the percentages of students whose scores fall into each interval are estimated. If the distributions have a similar shape, then the percentages in a given 10-point interval will be similar in both scales, and the linking is considered to have produced comparable scales. In the three assessments since this procedure has been followed, the distributions for the two components have proven very similar.

Other checks on comparability are also carried out. For a more detailed explanation of the linking process, see Allen, et al. (1995).

QUESTION 17:

What types of reports does NAEP produce? What reports are planned for the 1996 assessment?

ANSWER:

NAEP is an information system designed to provide various national and local audiences with results that will help them monitor and improve the educational system. To have maximum utility, NAEP reports must be clear and concise and must be delivered in a timely fashion.

NAEP has produced a comprehensive set of reports for the 1994 assessments in reading, U.S. history, and geography. These reports address the needs of national and state policy makers, educators, parents, and concerned citizens.

To meet the challenges associated with reporting the 1996 assessment results, a multifaceted approach has been designed in which a range of different publications will be produced. This strategy both introduces innovative elements into NAEP reporting and maintains aspects of the NAEP program that have proved successful in the past.

Key to this reporting strategy is the notion that, given the wide range of data available, all NAEP reports should be targeted to specific audiences rather than being directed to meeting an "average" need for information. The audiences interested in NAEP results include parents, teachers, school administrators, legislators, and researchers; targeting each report to a subset of the interested audiences should increase the impact and appeal of the NAEP reports. In addition, selected NAEP reports are now available electronically on the World Wide Web (http://www.ed.gov/NCES/naep), making them more generally accessible.

FURTHER DETAILS:

NAEP reports for different audiences

Building on the recent NAEP 1994 reports, NAEP proposes to continue the effort to produce reports that are technically sound and that address the needs of the many NAEP audiences. The NAEP 1994 reading results were first released in *NAEP 1994*

Reading: A First Look in April 1995 — one year following the completion of the data collection phase of the assessment. Similar reports were also produced for the NAEP 1994 U.S. history and geography assessments. The *First Look* reports represent an important step toward shorter, more readable reports that utilize bulleted text, simplified tables, and clear graphics. The response to these reports was positive and served as a springboard for the proposed 1996 reporting plans. The 1996 reports are in keeping with the National Center for Education Statistics (NCES) and National Assessment Governing Board (NAGB) commitment to provide valid NAEP results to interested audiences in a format that is accessible and on a time frame that is responsive.

The following reports are planned for the NAEP 1996 assessments.

As with the 1994 reports, plans are to place many of these reports on the NCES Web site (http://www.ed.gov/NCES/naep).

- ***NAEP Report Cards*** will be shorter than previous reports and will be directed to the needs of national and state policy makers.
- ***Snapshots*** will be concise, nontechnical four-page summaries of major findings reported in the *Report Cards*. The *Snapshots* will serve as an overview for policy makers, the press, and the public. In 1994, this summary appeared as *At A Glance*.
- ***Update Reports*** will be nontechnical reports that directly address questions of interest to parents, local school board members, and the concerned public.
- ***Instructional Reports,*** including many of the educational and instructional materials available from NAEP assessments, are intended for educators, school administrators, and subject-matter experts.
- ***State Reports.*** For each jurisdiction that participates in the NAEP 1996 state-level mathematics (grades 4 and 8) and science (grade 8) assessments, customized reports will be produced to highlight the results from that jurisdiction. Mathematics will be reported at the state level for the third time (1990, 1992, and 1996) and science will be reported at the state level for the first time. The *NAEP 1996 State Reports* will build on the computer-generated reporting system used successfully in 1990, 1992, and 1994. These reports, one for each participating state, are intended to be most helpful to

state policy makers, state departments of education, and chief state school officers. As in past state-level NAEP assessments, the state testing directors and the state NAEP coordinators will be active participants in the production of the NAEP 1996 State Reports.

- ***Cross-State Data Compendia,*** first produced for the state reading assessment in 1994, have an intended audience of researchers and state testing directors. They are viewed as reference documents to accompany other reports. The *Compendia* will report state-by-state results for variables *discussed in the Report Cards, State Reports, Update Reports, and Instructional* Reports.
- ***Trend Report*** describes patterns and changes in student achievement as measured through the long-term trend assessments in mathematics, science, reading, and writing. The report will present trends for the nation and for selected demographic subgroups (based on race/ethnicity, gender, region, parents' level of education, and type of school).
- ***Focused Reports,*** as with past NAEP assessments, will be developed to explore in-depth questions with broad educational implications. *Focused Reports* will yield information of interest to educators, policy makers, and psychometricians, as well as interested citizens.
- ***Almanacs*** present extensive tabular summaries of the cognitive and background data from the mathematics and science assessments at the state and national levels. They serve as archival records and are made available by NCES to state assessment directors and researchers. The *NAEP 1996 Almanacs* will be made available on CD-ROM, with easy-to-use NAEP Almanac Viewer software.
- ***Technical Reports*** document in detail all aspects of a given national or state assessment, including the sample design, instrument development, data collection, and analysis procedures. The technical reports do not provide the results of the assessment, but rather information on how those results were derived. Three technical reports are planned for the 1996 NAEP — one for the state mathematics assessment, one for the state science assessment, and one for the full national assessment.

In addition to the wide range of reports, NAEP will provide continued service to states and local school districts to help them better understand and utilize the results from NAEP assessments. The dissemination and use of NAEP results will be examined in an attempt to broaden the use of NAEP reports. One exciting avenue implemented recently is the electronic distribution of NAEP 1994 reports and almanacs through the World Wide Web (http://www.ed.gov/NCES/naep). The use of the Web brings NAEP onto the "information superhighway" and possibly expands the number of readers for NAEP reports.

QUESTION 18:

In what ways can educators apply NAEP resources to their work? How can they use frameworks, questions, and reports?

ANSWER:

NAEP materials such as frameworks, released questions, and reports have many uses in the educational community. For instance, NAEP frameworks have been considered during revision of a state's curriculum; released constructed-response questions with their scoring guides have provided models of innovative assessment practices.

NAEP findings are reported in many publications specially targeted to educators. In addition, NAEP staff host seminars for discussion of NAEP results and their implications.

FURTHER DETAILS:

NAEP frameworks as resources

NAEP frameworks are useful to instructors and curriculum planners. Frameworks present and explain what experts in a particular subject area consider important. Each framework outlines the subject, often with examples, in ways that may present new perspectives to teachers. Frameworks frequently provide theoretical information on problem-solving through their descriptive classification of cognitive levels (e.g., in the mathematics framework, conceptual understanding, procedural knowledge, and problem solving). Educators can consider how these cognitive levels relate to the various subject content areas, and can evaluate to what degree classroom instruction and assessment focus on each of the three cognitive levels. For example, a certain instructor may study the framework for a particular field and find that most of his or her instruction addresses only procedural knowledge; the instructor can then include more problems at a higher cognitive level, perhaps prompted by examples in the framework.

Several states have used the frameworks publications in reviewing their NAEP results and in developing recommendations for teachers, and NAEP staff members who

are conversant with particular frameworks have been invited to lend their expertise to state curriculum committees.

Uses of released NAEP questions

After each assessment, NCES releases nearly one-third of the questions, and copies are made available to the interested public. (For more information on released items, see Question 5.) These questions come with answer keys, content and process descriptions, and information on the percentages of students answering the questions correctly. These released questions have often served as models for teachers who wish to develop their own classroom assessments. One school district used released NAEP reading questions to construct its own district-wide test. Another school district has used scoring guides for released NAEP reading questions to instruct its teachers on scoring-guide construction.

Uses of NAEP reports

NAEP reports such as the focus report on mathematical problem-solving have proven useful to teachers. Other focus reports are planned for 1996, as described in Question 17. Further, NAEP staff have conducted NAEP seminars for school districts across the country in which NAEP results and their implications at the local level are discussed. In 1996, NCES began putting NAEP reports and almanacs on their World Wide Web site for viewing, printing, and downloading the data; this should increase the utility of NAEP results by making them even more accessible.

QUESTION 19:

How can NAEP data and assessment results be used for further exploration of education and policy issues? What kinds of technical assistance does NAEP provide?

ANSWER:

Members of the educational research community may obtain permission from the National Center for Education Statistics (NCES) to use NAEP data. Educational Testing Service (ETS) provides technical assistance, either as a public service or under contract, in the use of these data.

NAEP results are provided in formats designed for ease of access by the public. Reports have been tailored to specific audiences (see Question 17) and are widely disseminated. Beginning with the 1994 assessments, reports and almanacs have been placed on the World Wide Web to provide even easier access for the public (http://www.ed.gov/NCES/naep).

FURTHER DETAILS:

NAEP data

Because of its large scale, the regularity with which it is administered, and the rigid quality control on its data collection and analysis, NAEP provides a wealth of opportunities for secondary data analysis. NAEP data are used by researchers with many interests — from educators with policy questions to cognitive scientists studying the development of abilities across the three grades assessed by NAEP.

World Wide Web presence. In 1996 NCES began putting NAEP reports and almanacs on their World Wide Web site (http://www.ed.gov/NCES/naep) for viewing, printing, and downloading of data.

Software and data products. NAEP has developed products to support both internal and external dissemination of NAEP results and data to many different analysis audiences. ETS began developing the data products for the 1990 NAEP, with new capabilities and refinements added in 1992, 1994, and 1996. The data products for 1996

include the secondary-use version of the NAEP database on CD-ROM, comprehensive user guides, selection and extraction software, and analysis modules. In 1996 several programs were released for the Windows system. For instance, there are now Windows versions of

- *the NAEP Almanac Viewer* for searching and displaying crosstabulated variable tables,
- *the NAEP Data Extraction Program, NAEPEX,* for choosing variables, extracting data, and generating SAS or SPSS control statements, and
- *the NAEP SPSS Modules* for performing crosstabulations and regression analyses.

Using these products, ETS and NCES will continue to conduct workshops that promote secondary analysis of NAEP data.

NAEP technical assistance

Seminars. Once a year, a four-day seminar is offered: the NCES Advanced Studies Seminar on the Use of NAEP Database for Research and Policy Discussion. The purpose of this seminar is to stimulate interest in using NAEP data to address educational research questions, to enhance participants' understanding of the methodological and technological issues relevant to NAEP and to demonstrate the steps necessary to accurately conduct statistical analyses of the NAEP data. The seminar includes both formal and hands-on instruction and provides participants the opportunity to learn and work with currently available software packages specifically designed for NAEP analysis.

Special analyses. Under the cooperative agreement with NCES, ETS develops software to assist secondary users of NAEP data and regularly provides technical assistance to these researchers, either by providing them with information about data characteristics or by contracting to run analyses for them. Because requests for technical assistance occur regularly, NAEP staff have developed a procedure for expediting such requests.

Special studies. NAEP is also responsive to requests for empirical investigation of educational issues. Over the years, nearly every assessment has contained a special

study that has grown out of the need for information expressed by the public and by NCES. For instance, because of the desire to compare our students with the students assessed in the Third International Mathematics and Science Survey (TIMSS) of 1995, special studies were planned. These studies were supported by the special advanced targeted assessments in mathematics (administered to students of advanced mathematics in grades 8 and 12) and in science (administered to students of advanced science in grade 12) that were part of the NAEP assessment in 1996.

Forum for discussion of educational issues and policy

At the state level, the NAEP NETWORK meetings provide a forum for educational concerns in which testing directors and NAEP coordinators in individual states, territories, and other jurisdictions as well as representatives from nonpublic school organizations and associations may participate. The NETWORK also offers information about upcoming assessments and an opportunity for input from those involved in the state-level NAEP assessment.

QUESTION 20:

Can NAEP results be linked to international assessment data?

ANSWER:

NAEP results from the state assessment are regularly linked to results from the national assessment. Development of these linking procedures has given impetus for linking NAEP with other non-NAEP assessments. The success of linking NAEP 1992 results with those from the International Assessment of Educational Progress of 1991 (IAEP) has encouraged plans to link the NAEP 1996 results from the assessments in mathematics and science to those from the Third International Mathematics and Science Study (TIMSS) of 1995.

FURTHER DETAILS:

Pashley and Phillips (1993) investigated a linking of mathematics performance in the 1991 IAEP and the 1992 NAEP. Sample data to allow such a linking were collected in 1992 from U. S. students who were administered both instruments. One state, Colorado, drew a large enough sample to compare itself to all 20 participating countries.

The relation between mathematics proficiency in the two assessments was modeled (by regression analysis), and the model was used as the basis for projecting IAEP scores from non-U.S. countries onto the NAEP scale.

The study's authors considered their results very encouraging; the relation between the IAEP and NAEP assessments was quite strong and could be modeled well. However, as the authors point out, the results should be considered in the context of the similarity of construction and scoring for the two assessments. They advise that other studies be initiated cautiously, even though the path to linking assessments is now better understood.

In 1989 the United States made known its interest in international comparisons, especially in mathematics and science. The National Education Summit convened that year and adopted goals for education. Goal 4 states that by the year 2000, "U.S. students will be the first in the world in science and mathematics achievement." Since that

pronouncement, a variety of approaches have been suggested for collecting data that could help monitor progress toward that goal.

One of the most ambitious efforts was afforded by the 1995 TIMSS. The data from this study will be available at about the same time as the NAEP data for the 1996 mathematics and science assessments. It will be a rewarding exercise to link the results from the 1995 TIMSS and the 1996 NAEP. Through this linking of results from U.S. students with those of their academic peers in about 50 other countries, it is hoped that our progress toward National Education Goal 4 will be clear.

QUESTION 21:

Who evaluates and validates NAEP?

ANSWER:

The "Nation's Report Card" and its findings have a considerable impact on the public's understanding of student academic achievement. Because NAEP has such a unique and prominent role, care must be taken to ensure that its findings are valid and reliable. Realizing this, Congress has consistently passed legislation that established panels both to study the validity of NAEP and to evaluate the assessment as a whole. The National Center for Education Statistics (NCES) has responded to this mandate by establishing a variety of expert panels to study NAEP. To date, these panels have produced a series of reports addressing a number of critical NAEP issues.

FURTHER DETAILS:

By law, the Commissioner of NCES is responsible for providing "continuing reviews of the National Assessment, including validation studies . . ." (P.L. 100-297, Sec. 3403 [I] [9] [A]). In an effort to fulfill this mandate, a contract was awarded to the Center on Evaluation, Standards, and Student Testing (CRESST) at the University of California at Los Angeles (UCLA) in conjunction with the University of Colorado at Boulder and RAND in 1989 to establish a Technical Review Panel (TRP). Starting in 1989, the TRP produced a series of studies on specific questions relating to the validity of the interpretations of NAEP

results. These studies included:

- the quality of NAEP data;
- the number and character of NAEP scales;
- the robustness of NAEP trend lines;
- the trustworthiness of an interpretation of group comparisons;
- the validity of interpretations of NAEP anchor points and achievement levels;
- the linking of other test results to NAEP;
- the effects of student motivation on performance;
- the adequacy of NAEP data on student background and instructional experiences; and
- what is understood from NAEP reports by educators and policy makers.

At the end of the project period, the TRP completed its work on NAEP and produced a report summarizing the results of their studies (Linn, Koretz, & Baker, 1996).

On October 20, 1994, NAEP was reauthorized when the President signed Public Law 103-382. Similar to the earlier law, this statute mandated that "the Secretary shall provide for continuing review of the National Assessment, State Assessments, and student performance levels by one or more nationally recognized organizations. . ." NCES responded to this legislation by funding a contract for the establishment of the NAEP Validation Studies (NVS) Panel. The NVS Panel was formed in October, 1995, and is currently working on a number of validity studies. Two reports are scheduled for release in 1996: Optimizing State NAEP and The Impact of Accommodations. Other reports to be released later include a study on sampling small populations, the information possibilities of performance (i.e., constructed response) questions, and equating error in NAEP.

Both the Technical Review Panel and the NAEP Validity Studies Panel were designed to study the validity of the NAEP national assessments. A third panel was commissioned to study the validity of the state component of NAEP. In the spring of 1988, Congress enacted Public Law 100-297 authorizing a NAEP Trial State Assessment (TSA). In authorizing the TSA, Congress called for an independent evaluation of "the

feasibility and validity of state assessments and the fairness and accuracy of the data they produce." Responding to the legislation, NCES funded a grant to National Academy of Education (NAE) Panel on the Evaluation of the NAEP Trial State Assessment Project in 1990. Between 1992 and 1996 the NAE panel produced a host of reports evaluating the validity of the TSA. The Panel recently completed its evaluation of the 1994 state NAEP assessment and will be releasing a capstone report in November 1996. This last report draws conclusions about state NAEP and makes recommendations for future NAEP assessments. The NAE panel also evaluated the National Assessment Governing Board's performance standards.

Starting in 1996, the National Academy of Science (NAS) was awarded the contract to continue the evaluation of NAEP, both national and state. Toward that end, the NAS convened a panel of distinguished educators and other experts to carry out the evaluation activities as described by the Congressional mandate of 1994. NAS will be conducting a series of workshops, meetings, and to the extent necessary, empirical investigations and letter reports. The final report to congress will provide their concluding observations and recommendations.

BIBLIOGRAPHY

Allen, N. L., Mazzeo, J., Ip, E. H. S., Swinton, S., Isham, S. P., & Worthington, L. (1995). Data Analysis and scaling for the 1994 Trial State Assessment in reading. In J. Mazzeo, N. L. Allen, & D. L. Kline, *Technical report of the NAEP 1994 Trial State Assessment in reading* (pp. 169-219). Washington, DC: National Center for Education Statistics.

Allen, N. L., Kline, D. L., & Zelenak, C. A. (1996). *The NAEP 1994 Technical Report.* Washington, DC: National Center for Education Statistics.

Beaton, A. E., & Zwick, R. (1990). *The effect of changes in the National Assessment: Disentangling the NAEP 1985-1986 reading anomaly.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beatty, A. S., Reese, C. M., Persky, H. R., & Carr, P. (1996). *NAEP 1994 U.S. history report card: Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

Campbell, J. R., Reese, C. M., O'Sullivan, C., & Dossey, J. A. (1996). *NAEP 1994 Trends in Academic Progress: Achievement of U.S. students in science, 1969 to 1994; mathematics, 1973 to 1994; reading, 1971 to 1994; and writing, 1984 to 1994.* Washington, DC: National Center for Education Statistics.

College Board (1994). *Mathematics framework for the 1996 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

College Board (1994). *Science framework for the 1996 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Council of Chief State School Officers (1992). *Geography framework for the 1994 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Council of Chief State School Officers (1992). *Reading framework for the 1992 and 1994 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Council of Chief State School Officers (1992). *U.S. history framework for the 1994 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Green, J. L., Burke, J., & Rust, K. L. (1995). Sample design and selection. In J. Mazzeo, N. Allen, & D. Kline, *Technical report of the NAEP 1994 Trial State Assessment in reading* (pp. 35-69). Washington, DC: National Center for Education Statistics.

Linn, R. L., Koretz, D., & Baker, E. L. (1996). Assessing the validity of the National Assessment of Educational Progress: NAEP Technical Review Panel white paper. Washington, DC: U.S. Department of Education.

Mullis, I., Dossey, J. A., Campbell, J. R., Gentile, C. A., O'Sullivan, C., and Latham, A. S. (1994). *NAEP 1992 trends in academic progress: Achievement of U.S. students in science, 1969 to 1992; mathematics, 1973 to 1992; reading, 1971 to 1992; writing, 1984 to 1992.* Washington, DC: National Center for Education Statistics.

National Assessment of Educational Progress (1986). *Math Objectives, 1985-86 Assessment.* Washington, DC: U.S. Department of Education.

National Assessment of Educational Progress (1984). Reading Objectives, 1983-84 Assessment. Washington, DC: U.S. Department of Education.

National Assessment of Educational Progress (1986). *Science Objectives: 1985-1986 Assessment.* Washington, DC: U.S. Department of Education.

National Assessment of Educational Progress (1989). *Science Objectives: 1990 Assessment.* Washington, DC: U.S. Department of Education.

National Assessment of Educational Progress (1982). Writing Objectives: 1984 Assessment. Washington, DC: U.S. Department of Education.

National Assessment of Educational Progress (1987). *Writing Objectives: 1988 Assessment.* Washington, DC: U.S. Department of Education.

Pashley, P. J., & Phillips, G. W. (1993). *Toward world-class standards: A research study linking international and national assessments.* Princeton, NJ: Educational Testing Service.

Persky, H. R., Reese, C. M., O'Sullivan, C. Y., Lazer, S., Moore, J., & Shakrani, S. (1996). *NAEP 1994 geography report card : Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

Wallace, L., & Rust, K. F. (1996). Sample Design. In N. Allen, D. Kline, & C. Zelenak, *NAEP 1994 technical report* (Chapter 3). Washington, DC: National Center for Education Statistics.

Williams, P. L., Lazer, S., Reese, C. M., & Carr, P. (1995). *NAEP 1994 U.S. history: A first look — Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

Williams, P. L., Reese, C. M., Campbell, J. R., Mazzeo, J., & Phillips, G. W. (1995). *NAEP 1994 reading: A first look — Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

Williams, P. L., Reese, C. M., Lazer, S., & Shakrani, S. (1995). *NAEP 1994 geography: A first look — Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

FURTHER READING

Gentile, C., Martin-Rehrmann, J., & Kennedy, J. H. (1995). *Windows into the classroom: NAEP's 1992 writing portfolio study.* Washington, DC: U.S. Department of Education.

Lapointe, A. E., Askew, J., & Mead, N. A. (1992). *Learning mathematics.* Princeton, NJ: Educational Testing Service.

Lapointe, A. E., Askew, J., & Mead, N. A. (1992). *Learning science.* Princeton, NJ: Educational Testing Service.

Lazer, S. (1992). *Learning about the world.* Princeton, NJ: Educational Testing Service.

Lindquist, M. M., Dossey, J. A., & Mullis, I. (1995). *Reaching standards: A progress report on mathematics. A policy information perspective.* Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1992). *Linking Educational Assessments.* Princeton, NJ: Educational Testing Service.

Modell, J. (1993). *Desire to learn: A comparative view of schooling in children's lives.* Princeton, NJ: Educational Testing Service.

Mullis, I., Jenkins, F. L., & Johnson, G. (1994). *Effective schools in mathematics.* Washington, DC: National Center for Education Statistics.

*U.S. Government Printing Office: 1996 — 420-699/60363

Subject Areas Assessed by NAEP

Year	NATIONAL NAEP: Main Assessment	NATIONAL NAEP: Long-Term Trend	STATE NAEP
1998	Civics, Reading, Writing		Reading Grades 4, 8 Writing Grade 8
1997	Arts (Grade 8)		
1996	Mathematics, Science	Mathematics, Science, Reading, Writing	Mathematics Grades 4, 8 Science Grade 8*
1994	Geography, U.S. History, Reading	Mathematics, Science, Reading, Writing	Reading[3] Grade 4
1992	Mathematics, Reading, Writing	Mathematics, Science, Reading, Writing	Mathematics[3] Grades 4, 8 Reading[3] Grade 4
1990	Mathematics, Science, Reading	Mathematics, Science, Reading, Writing	Mathematics[3] Grade 8
1988	Civics, Document Literacy[2], Geography[2], U.S. History, Reading, Writing	Civics[1], Mathematics, Science, Reading, Writing	*Since 1994, states have assessed nonpublic schools as well as public schools.*
1986	Computer Competence, U.S. History[2], Literature[2], Mathematics, Science, Reading	Mathematics, Science, Reading[4]	*Department of Defense schools were assessed at Grades 4 and 8.
1984	Reading, Writing	Reading, Writing	
1981-82[5]	Mathematics, Science, Citizenship / Social Studies[1]	Mathematics[1], Science[1]	
1979-80	Reading / Literature, Art	Reading[1]	
1978-79	Art, Music, Writing		
1977-78	Consumer Skills[2], Mathematics	Mathematics[1]	
1976-77	Basic Life Skills[2], Science	Science[1]	
1975-76	Citizenship / Social Studies, Mathematics[2]	Citizenship / Social Studies[1]	
1974-75	Art, Index of Basic Skills, Reading	Reading[1]	
1973-74	Career and Occupational Development, Writing		
1972-73	Mathematics, Science	Mathematics[1], Science[1]	
1971-72	Music, Social Studies		
1970-71	Literature, Reading	Reading[1]	
1969-70	Citizenship, Science, Writing	Science[1]	

[1] This assessment appears in reports as part of long-term trend. Note that the Civics assessment in 1988 is the third point in trend for Citizenship / Social Studies in 1981-1982 and in 1975-1976. There are no points on the trend line for Writing before 1984.

[2] This was a small, special-interest assessment administered to limited national samples at specific grades or ages and was not part of a main assessment. Note that this chart includes only assessments administered to in-school samples; not shown are several special NAEP assessments of adults.

[3] State assessments in 1990-1994 were referred to as Trial State Assessments (TSAs).

[4] The 1986 long-term trend Reading assessment is not included on the trend line in reports because the results for this assessment were unusual. Further information on this reading anomaly is available in Beaton and Zwick (1990).

[5] Explanation of format for Year column: Before 1984, the main NAEP assessments were administered in the fall of one year through the spring of the next. Beginning with 1984, the main NAEP was administered after the new year in the winter, although the assessments to measure long-term trend continued with their traditional administration in fall, winter, and spring. Because the main assessment is the largest component of NAEP, beginning with 1984 we have listed its administration year rather than the two years over which the trend continued to be administered. Note also that the state component is administered at essentially the same time as the main NAEP.

BEST COPY AVAILABLE

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (A)

NCES 97–586

BEST COPY AVAILABLE